

UCLA

UCLA Previously Published Works

Title

A clinically applicable integrative molecular classification of meningiomas

Permalink

<https://escholarship.org/uc/item/7b46s9w7>

Journal

Nature, 597(7874)

ISSN

0028-0836

Authors

Nassiri, Farshad

Liu, Jeff

Patil, Vikas

et al.

Publication Date

2021-09-02

DOI

10.1038/s41586-021-03850-3

Peer reviewed



Published in final edited form as:

Nature. 2021 September ; 597(7874): 119–125. doi:10.1038/s41586-021-03850-3.

A Clinically Applicable Integrative Molecular Classification of Meningiomas

Farshad Nassiri, MD^{1,2,12}, Jeff Liu, PhD^{1,12,*}, Vikas Patil, PhD^{1,12,*}, Yasin Mamatjan, PhD^{1,12,*}, Justin Z. Wang, MD^{1,2,12,*}, Rupert Hugh-White, PhD³, Andrew Macklin, MSc⁴, Shahbaz Khan, PhD⁴, Olivia Singh, MSc¹, Shirin Karimi, MD¹, Rosario I. Corona, PhD³, Lydia Y Liu, PhD^{3,11}, Caroline Y Chen, MD³, Ankur Chakravarthy⁴, Qingxia Wei, PhD¹, Bharati Mehani, PhD⁵, Suganth Suppiah, MD^{1,2,12}, Andrew Gao, MD⁶, Adriana M. Workewych, BSc¹, Ghazaleh Tabatabai, MD^{7,12}, Paul C. Boutros, PhD^{3,8,12}, Gary D. Bader, PhD^{9,10}, Daniel D. de Carvalho, PhD^{4,11}, Thomas Kislinger, PhD^{4,11,12}, Kenneth Aldape, MD^{1,5,12,**}, Gelareh Zadeh, MD, PhD^{1,2,12,**}

¹MacFeeters Hamilton Neuro-Oncology Program, Princess Margaret Cancer Centre, University Health Network and University of Toronto, ON, Canada.

²Division of Neurosurgery, Department of Surgery, University of Toronto, Toronto, ON, Canada

³Jonsson Comprehensive Cancer Center, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.

⁴Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada.

⁵Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA.

⁶Laboratory Medicine Program, University Health Network, Toronto, Ontario, Canada

⁷Department of Neurology & Interdisciplinary Neuro-Oncology, Hertie Institute for Clinical Brain Research, Center for Neuro-Oncology, Comprehensive Cancer Center, University Hospital Tübingen, Tübingen, Germany

⁸Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA

⁹The Donnelly Center, University of Toronto, Toronto, ON, Canada

¹⁰Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

Correspondence should be addressed to: Gelareh Zadeh, MD, PhD, Division of Neurosurgery, University of Toronto, MacFeeters Hamilton Neuro-Oncology Program, Princess Margaret Cancer Centre, University, Health Network, Toronto, Ontario, Canada, gelareh.zadeh@uhn.ca.

*These authors contributed equally

**These authors contributed equally

Author Contributions:

F.N., K.A., G.Z. conceived and designed the study. F.N. and J.W. collected all the biomaterials and clinical data. K.A., A.G., and S.K. reviewed biomaterials. F.N. prepared specimens for whole-exome sequencing, DNA methylation, mRNA sequencing and single cell RNA sequencing. S.K. and A.M. carried out proteomic experimentation. O.S., J.W., and F.N. carried out immunohistochemical experimentation and analyses. F.N. and J.W. carried out in-vitro and in-vivo experimentation. F.N., J.L., Y.M., V.P. A.C., R.H.W., R.I.C., L.Y.L., C.Y.C. and B.M. contributed to the data processing and analyses. F.N., K.A., P.C.B., G.B., D.D.D.C., T.K., and G.Z. contributed to data interpretation. F.N. and A.M.W. organized the Figures. F.N. and G.Z. wrote the first draft as well as subsequent revisions and response to reviewers. All authors contributed to final data interpretation and critical revision of the manuscript and approved the final version of the manuscript. G.Z. supervised all aspects of the study.

¹¹Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

¹²The International Consortium on Meningiomas

Abstract

Meningiomas are the most common primary intracranial tumor in adults¹. Symptomatic patients are treated with surgery and there are no effective medical therapies. While the World Health Organization (WHO) histopathological grade of the tumor and the extent of resection at surgery (Simpson grade) are associated with recurrence of disease, they do not accurately reflect the clinical behavior of all meningiomas². Molecular classifications of meningioma that reliably reflect tumor behaviour and inform on therapies are greatly needed. Here we introduce four novel consensus molecular groups of meningioma by combining DNA somatic copy number aberrations, DNA somatic point mutations, DNA methylation and messenger RNA abundance in a unified analysis. These molecular groups more accurately predicted clinical outcomes in comparison to existing classification schemes. Each molecular group showed distinctive and prototypical biology (immunogenic, *NF2*-wildtype, hypermetabolic, and proliferative) that informed on novel therapeutic options. Proteogenomic characterization reinforced the robustness of the newly defined molecular groups and uncovered highly abundant and group-specific protein targets that we validated using immunohistochemistry. Single cell RNA sequencing revealed inter-individual variations in meningioma as well as variations in intrinsic expression programs in neoplastic cells that mirrored the biology of the molecular groups we have identified.

While previous studies on meningioma have provided important insights into the possibility for molecular data to refine meningioma classification^{3–8}, formal integration of multiple molecular datatypes in a unified analysis has not been performed. Here, we have assembled a large cohort of meningiomas that were enriched for the uncommon higher-grade tumors with matched multidimensional molecular and high quality clinical data. We generated matched DNA somatic copy number, DNA point mutation, DNA methylation, transcriptome, and proteomic data to create a TCGA-style resource for meningiomas that we supplemented with single cell RNA sequencing data. By integrating multiple datatypes in a unified analysis as done in other cancers^{9–12}, we define a novel molecular taxonomy for meningiomas with direct clinical relevance.

Patient samples and clinical data

We used meningioma samples from 121 patients to define molecular groups and 80 samples from an independent cohort to assess generalizability. Samples were selected based on availability of clinical data as well as quality and quantity of tissue for analyses. Our cohort reflects the real-life diversity of patients with meningiomas and includes a substantial number of WHO grade 2 and 3 meningiomas, which have been understudied to date due to their rarity. We performed whole exome sequencing for germline polymorphisms, somatic point mutations and somatic copy number alterations; EPIC array profiling for DNA methylome analysis; and messenger RNA-sequencing for transcriptome analysis on all 121 tumors in the discovery cohort, with whole-cell proteomics performed on 96 of these (Fig. 1a). DNA methylation was also performed on five healthy meninges for methylome

comparisons. Eight of these tumors and two healthy meninges samples were profiled by single nuclear RNA sequencing to examine intratumoral heterogeneity. Grading was confirmed by two independent neuropathologists in accordance with the most recent 2016 WHO classification criteria. All samples were annotated with detailed high-quality clinical data elements that were established *a priori* (see Methods and Supplementary Table 1).

Interdependencies of datatypes

To examine relationships between datatypes, we computed the Mutual Information (MI) metric for each gene between all pairwise combinations of datatypes and compared this permuted null distribution¹³. MI values of zero indicate orthogonal information. We found that the distribution of MI were statistically significantly different between different datatype comparisons (Extended Data Fig. 1a). Moreover, consensus clustering of normalized MI values for genes where MI was significant for at least one datatype pair revealed four gene clusters, each defined by distinct patterns of dependence between datatypes at different levels of the central dogma, pointing to the potential value for formal unsupervised integration of multiple datatypes.

Multiplatform integrative analyses

We next sought to combine whole-exome sequencing and copy number, DNA methylation, and mRNA sequencing data using the Cluster of Cluster Algorithm (COCA)^{9–12}. With this approach, cluster assignments from individual platform analyses are subjected to additional (second-order) clustering to examine higher-order relationship of samples across molecular features.

Unsupervised sample-wise clustering of gene level somatic copy number alterations, DNA methylome and transcriptome data in isolation revealed six stable subgroups for each datatype with clinically-relevant and significant differences in outcome (Fig 1b, Extended Data Fig. 1b,d,f). Cluster assignments across datatypes were neither identical nor orthogonal (Fig. 1c) and cluster associations with outcome were unique for each datatype (Extended Data Fig. 1c,e,g).

COCA combining six copy number clusters with six DNA methylation and six mRNA abundance clusters converged to reveal four novel stable molecular groups (MG 1–4) of meningioma (Fig. 1d and Extended Data Fig. 1h). RNA cluster assignments were strongly associated with MG1, MG3, and MG4 while CNA and DNA methylation cluster assignments were most strongly associated with MG2, and the relative importance of these datatypes were confirmed by formal unsupervised integration of two datatypes at a time (Supplementary Table 2). Tumors spanning all WHO grades were represented in each molecular group, with the exception of MG1 that was composed of only WHO grade 1 and 2 tumors. Higher WHO grade tumors were enriched in MG3 and MG4 (Fisher's Exact test, $P=5.49\times 10^{-7}$). Importantly, a clear one-to-one relationship of molecular group to WHO grade was not evident (Extended Data Fig. 1i), prompting us to examine the clinical relevance of these newly defined integrative molecular groups.

Clinical relevance of integrative MGs

While the discovery of the four MGs in this study was agnostic to patient outcomes, these groups were characterized by distinct and divergent patterns of recurrence-free survival (Fig. 1f). Overall, patients with MG3 and MG4 tumors had statistically shorter times to recurrence (Log Rank Test, $P=5\times 10^{-15}$), with the most unfavorable outcomes in MG4 tumors. MG classification was independently associated with recurrence-free survival even after accounting for known prognostic clinical factors including WHO grade, extent of surgical resection, and receipt of adjuvant radiotherapy by Multivariable Cox regression (see Supplementary Table 3). Significant differences in recurrence patterns persisted across MGs when tumors were analyzed separately for each WHO grade (Extended Data Fig. 1j–l). Classification by MG groups was superior in predicting time to recurrence compared to the WHO grade, previously described methylation-based classifications³ as well as classification by cluster assignments from each datatype individually (Fig. 1f). We confirmed the generalizability of MG classification and outcomes in an independent cohort using mRNA signatures (Extended Data Fig. 2). This framework provides a blueprint for future independent validation and ongoing assessment of generalizability.

Mutational profiles of MGs

We next examined the somatic point mutational profiles of MG groups. While *NF2* was predictably the most commonly point mutated gene, its prevalence differed significantly across MGs without distinct positional bias (Fig. 2a and Extended Data Fig. 3a). Nearly all MG1 meningiomas were *NF2*-mutated, whereas *NF2* mutations were extremely rare in the MG2 tumors (88% vs 9%, Fisher's Exact test, $P=5.9\times 10^{-8}$). Conversely, the previously described non-*NF2* mutations in *TRAF7*, *AKT1*, *KLF4* and *POLR2A* were exclusively identified in the MG2 tumors at frequencies of 25%, 13%, 13% and 6%, respectively (Fisher's Exact test, $P=1.20\times 10^{-8}$).

We identified novel, statistically significant, recurrent nonsynonymous somatic driver mutations in genes involved in chromatin modeling and epigenetic regulation (*KDM6A*, *CHD2*), as well as tumor suppressors (*PTEN*; Supplementary Table 4). Recurrent inactivating mutations in additional chromatin modeling (*CREBBP* $q=0.126$) and tumor suppressor genes (*FBXW7* $q=0.226$, *RBI* $q=0.250$) were also identified as subthreshold hits (Supplementary Table 4). These novel mutations that we discovered occurred at frequencies similar to other known meningioma driver genes (3–5%, Fig. 2a) and were collectively enriched in the aggressive phenotypes of meningioma distinguishing MG3 and 4 tumors from MG1 and 2 tumors (Fisher's Exact test, $P=0.002$). MG4 tumors had significantly greater mutational burden compared to MG 1–3 tumors ($P=1.6\times 10^{-3}$ Kruskal Wallis test, Extended Data Fig. 3b). The majority of point mutations in meningioma were clonal, with only a small subset seen as late-evolving drivers. (Extended Data Fig 3c–e). The specificity of different mutations for distinct MG groups was particularly notable given that the generation of MG groups was independent of point mutations.

Genomic disruptions across MGs

We next investigated the pattern of genome-wide copy number alterations across molecular groups (Extended Data Fig. 4a). MG1 tumors were relatively diploid with the exception of uniform loss of chromosome 22q, which in combination with concurrent *NF2* point mutations, results in biallelic *NF2* inactivation. There were two subsets of MG2 tumors, one in which tumors were copy number neutral but harbored mutations in *TRAF7*, *AKT1*, *KLF4*, or *SMO*, and the other in which tumors did not harbor mutations but had consistent polysomies of chromosomes 5, 12, 13, 17 and 20. MG3 and MG4 meningiomas were high aneuploidy tumors with losses in chromosomes 22q (93% and 86%), 1p (77% and 89%), chromosomes 6q (30% and 38%), 14 (47% and 35%) and 18 (19% and 38%). MG4 meningiomas also showed gain of chromosome 1q, and loss of chromosome 10 that was uncommon in MG3 meningiomas (34% vs 2%, $P=2.9\times 10^{-4}$ Fisher's Exact test and 38% vs 14%, $P=0.025$, respectively, Fisher's Exact test). Some *NF2*-wildtype MG3 and MG4 tumors showed silencing of *NF2* expression that was not associated with changes in methylation of the *NF2* gene (Extended Data Fig. 4b–c). The degree of total genomic disruption, quantified as the percent genome altered, was higher in MG3 (median 16.9%) and MG4 meningiomas (median 19.5%) compared to MG1 (median 3.5%) and MG2 (median 9.6%) tumors ($P=5.2\times 10^{-6}$, Kruskal Wallis test). This was further supported by more frequent non-recurrent interchromosomal fusion events in MG3 and MG4 tumors compared to MG1 and 2 meningiomas (Extended Data Fig. 4d and Supplementary Table 5). Taken together, these data point to an increase in genomic instability in MG3 and MG4 tumors that have the most unfavorable outcomes.

Gene expression networks of MGs

We next investigated the gene expression pathways associated with each MG (Fig. 2b and Extended Data Fig. 5a). MG1 tumors showed greater immune infiltration and enrichment of pathways involved in immune regulation and signalling (Fig. 2b, **inset** and Extended Data Fig. 5b). By contrast, immune signatures were down-regulated in MG4 tumors, and instead these tumors showed enrichment for pathways involved in cell-cycle regulation, as well as several critical and complementary proliferation-associated transcription factor networks (e.g., *MYC*, *FOXM1*, *E2F* etc.) and protein complexes (e.g., *mTORC1*, *CDKs*, *kinesins* etc.). MG3 tumors were uniquely enriched for pathways that converged onto the metabolism of several macromolecules. Although we identified two subsets of MG2 tumors by mutations and copy number, the transcriptomes of these subsets were distinctly correlated (Extended Data Fig. 5c–d), and collectively enriched for vascular and angiogenic pathways (Fig. 2b). Consequently, we designated the MGs as immunogenic (MG1), benign *NF2*-wildtype (MG2), hypermetabolic (MG3) and proliferative (MG4). Of note, the association of MG groups with outcomes was independent of molecular proliferation signatures (Extended Data Fig. 5e, and Supplementary Table 6).

We next sought to determine if the distinct expression pathways could be exploited to identify novel medical therapies in meningiomas by mapping Food and Drug Association (FDA) approved drugs to target genes in our enrichment network. We found that Vorinostat, a histone de-acetylase inhibitor, targeted several critical pathways specifically upregulated in

proliferative (MG4) meningiomas (Fig. 2b). Treatment with Vorinostat selectively decreased the viability of only MG4-tumor-patient-derived cell lines in comparison to patient-derived cell lines of other MGs (Fig. 2c and Extended Data Fig. 6a–b). By contrast, treatment of the same cell lines with a comparable agent, 5-azacytidine, had no effect on cell viability. Vorinostat also attenuated tumor growth (Fig. 2d) and improved survival (Fig. 2e) in intracranial xenografts of patient-derived MG4 cell lines in comparison to control mice (Extended Data Fig. 6c–d). Overall, these findings suggest that molecular groups may differ in treatment sensitivity to Vorinostat, warranting further in-human investigation.

Proteogenomic characterization of MGs

Using a single shot liquid chromatography tandem mass spectrometry approach we quantified a total of 6,568 unique protein-groups in 96 tumors with somatic mutation, epigenome and transcriptome data in our cohort. Enrichment scores of gene-sets by mRNA and proteome data were highly correlated when comparing samples of similar MGs (Extended Data Fig. 7a–c). Functional inference using protein data alone converged on biological networks that were highly similar to those obtained by transcriptome data (Fig. 3a and Extended Data Fig. 7d). Specifically, immunogenic (MG1) tumors were enriched for proteins involved in immunoregulation, whereas hypermetabolic (MG3) meningiomas harboured enrichment of protein pathways converging on nucleotide and lipid metabolism and proliferative (MG4) meningiomas were enriched for protein gene sets that regulate cell cycle and cell proliferation.

We next compared the association of mRNA and protein abundance with outcomes. Overall, the associations of protein and gene abundance with outcome correlated well (Pearson's $\rho=0.49$, 95% CI 0.47 to 0.50, $P<2.2\times 10^{-16}$). In fact, concordance was 213 times more likely (OR = 213.17, 95% CI 113.74 to 422.26) than non-concordance amongst the 682 genes that were significantly associated with outcome by either mRNA or protein data (Fig. 3b). It is noteworthy that genes associated with worse outcomes in both datatypes were involved in both cell cycle (FDR = 3.98×10^{-7} , hypergeometric test) as well as metabolism by oxidative phosphorylation (FDR = 2.9×10^{-55} , hypergeometric test).

We then identified proteins that were highly enriched in each MG by proteomic data: S100B for MG1, SCGN for MG2, ACADL for MG3 and MCM2 for MG4 (Supplementary Table 7, see Methods). We validated the enrichment of these proteins in each group by immunohistochemistry in a cohort of tumors with known molecular group status in a blinded fashion. Unbiased, digital quantitation of each protein marker showed strong concordance between immunohistochemistry and proteomic data, with protein markers discriminating between MG groups well (Fig. 3c). These results lay the groundwork showing the potential for molecular group classifications to be adopted in conventional neuropathology laboratories upon further independent validation.

Methylation characteristics of MGs

We next examined for differences in genome-wide DNA methylation patterns between healthy meninges and meningiomas. We identified two sets of probes that differentiated

healthy meninges from meningiomas as a whole (Extended Data Fig 8a). In one set, probes were fully hypomethylated in healthy meninges and progressively gained methylation across MGs, while in the other set, probes were fully hypermethylated in healthy meninges, and progressively lost methylation across MGs. (Extended Data Fig 8b). These patterns were similar when examining previously defined regions of the genome that either gain or lose methylation as a function of mitotic age (i.e. epigenetic mitotic clocks, Extended Data Fig 8c)^{14–16}, pointing to the possibility that aberrant DNA methylation processes may be associated with the most aggressive MGs, although differences in cell type composition may also be a contributing factor. We next identified transcription factors enriched in each MG on the basis of hypomethylated enhancer regions within each group (Extended Data Fig. 8d), known transcription-factor binding site motifs, and correlations with gene expression¹⁷. Hypomethylation at enhancer regions was associated with transcription factors that aligned to the biology of each MG that we defined by gene and protein expression (Extended Data Fig 8e–f).

Single cell map of meningiomas

To investigate heterogeneity in meningiomas, we performed droplet-based single nuclear RNA sequencing (snRNA-seq) on eight tumors selected to span all molecular groups and WHO grades, along with two healthy meninges samples for comparisons.

In total, 54,393 high-quality and accurately genotyped single nuclei were analyzed, and 14 distinct clusters were identified (Fig. 4a–d and Supplementary Figs 1 and 2). Cells were assigned to cell types on the basis of consensus between expression-based clustering (Extended Data Fig. 9a), inference of CNAs (Extended Data Fig. 9b–c), and annotation by canonical markers (Extended Data Fig. 10a). The majority of cells in our data were neoplastic (69%), while 14% were immune cells (macrophages and T-cells), 10% were fibroblasts, and 6% were endothelial cells.

Non-neoplastic cells from different patients clustered together by cell type, whereas neoplastic cells clustered distinctly by patient, representing the inter-individual variability of meningiomas (Fig. 4a and Extended Data Fig. 10b, and Supplementary Table 8). When neoplastic cells were considered in isolation, the variability between cells of different tumors was much larger than the variability within tumors ($F=65,538$ $P<2.2\times 10^{-16}$, One-way ANOVA), and within the limits of differences in detection rates of genes between cells, the expression of neoplastic cells most closely resembled bulk molecular signatures of their tumor of origin (Extended Data Fig 10c). Cycling neoplastic cells were enriched in MG3 and MG4 tumors ($P=2.2\times 10^{-2}$ and $P=1.49\times 10^{-2}$, respectively, mixed-effects) while immune cells were enriched in MG1 tumors ($P=1.8\times 10^{-2}$, mixed-effects; Extended Data Fig 10d–e). Indeed, deconvolution of bulk RNA seq data using single cell RNA-seq signatures confirmed that macrophages were enriched in MG1 tumors with additional differences in cell composition across MGs and healthy meninges (Fig. 4e and Extended Data Fig. 10e).

Heterogeneity by single cell

We first looked for discrete patterns of variation by clustering gene expression profiles of single cells from each sample individually using two independent clustering algorithms (Seurat and DBSCAN). When considering all cells within a sample, MG1–3 tumors showed several discrete clusters that were largely explained by the abundance of stromal or immune cell types whereas MG4 tumors, that were predominantly composed of neoplastic cells, did not show distinct clusters (Fig. 4f). We then selected the neoplastic cells of each tumor for additional sub-clustering using the same algorithms to examine the neoplastic component of each tumor more carefully. Again, using both algorithms we found that most samples harbored one dominant cluster, and less commonly, a second minor cluster of neoplastic cells. Copy number profiles of neoplastic cells were in general similar to those observed on bulk analyses and again did not show substantial variability between cells (Extended Data Fig. 9b–c). These findings were in line with our results from clonality assessment of bulk mutation data (Extended Data Fig. 3c–e), highlighting the relative rarity of subclonal expansion in meningiomas.

We then looked to identify programs that were intrinsically expressed in neoplastic cells and shared between samples by non-negative matrix factorization (NMF). In total, we identified 24 of such programs across neoplastic cells of different samples that clustered to four ‘meta-programs’ based on the degree of similarity by shared genes between modules (Fig. 4g and Extended Data Fig. 11a). The metaprograms were highly similar to the biology of the integrative molecular groups we defined earlier, and the distributions of the activation of these programs across cells of different tumors reflected this (Extended Data Fig. 11b). The most prominent program was related to cell cycle (FDR = 3.13×10^{-32} , hypergeometric test), and this program was reflective of discrete patterns of variability in most tumors (Extended Data Fig. 11b–c). Other programs included cellular metabolism (FDR = 7.66×10^{-3} , hypergeometric test), TNF α -inflammatory (FDR = 5.99×10^{-13} , hypergeometric test), and a general mesenchymal program (FDR = 2.12×10^{-15} , hypergeometric test) which generally showed more continuous patterns of variation (Extended Data Fig. 11c–d). Overall, these programs represent more subtle patterns of variation in meningiomas, however, the similarity of these programs, which are intrinsic to neoplastic cells, to the biology that we defined for the molecular groups we introduced in this study, points to the importance of these processes in meningioma biology. Indeed, deconvolution and partitioning of our bulk mRNA data using neoplastic and non-neoplastic signatures derived from our single cell RNA seq data showed a high degree of similarity to the molecular groups we define in this study (Extended Data Fig. 10g).

Conclusions

Here we present a key resource for the meningioma community with matched multidimensional bulk and single cell molecular and high-quality clinical data. By integrating multiple datatypes in a unified analysis, we define a novel molecular taxonomy for meningiomas (Extended Data Fig. 12) that supersedes existing molecular and clinically used classifications with the potential to inform on future iterations of recognized grading schemes.

Methods

Patient samples and clinical annotation

In total, the International Consortium on Meningiomas has accrued 670 fresh-frozen or paraffin-embedded meningioma samples for molecular analyses. Clinical data was collected for each sample using pre-established common data elements (CDEs) designed for reporting on molecular studies of meningioma. Definitions for CDEs were agreed upon using a systematic process of discovery, internal validation, external validation, and distribution. A total of 19 core CDEs (including age, sex, country of care, history of neurofibromatosis, history of malignancy, prior exposure of cranial radiation or chemotherapy, history of multiple meningiomas, timing of surgery, location of tumor, extent of resection at surgery, histopathological grade [WHO] and year of WHO classification system, recurrence status, time to recurrence from index surgery, prior irradiation to meningioma, time to last follow-up) were collected for all samples and an additional 14 supplemental CDEs (including race/ethnicity, hispanic race, diagnosis of meningioma syndrome, tumor size, Simpson grade, performance status at recurrence or last follow-up, second intervention for recurrence, time to second intervention, histopathological subtype of recurrent tumor, vital status, cause of death, time to death) were collected per sample, where possible. Collection of samples and clinical data was carried out in accordance with individual institutional ethics and review board guidelines.

For the present study focusing on integration of multiplatform molecular studies, tissue and blood samples were selected based on sufficient availability of specimens (>500mg tissue and >1ml blood/plasma). In total, 124 fresh-frozen meningioma samples and 5 healthy meninges samples from patients were collected for molecular analyses from the University Health Network Brain Tumor BioBank (Toronto) under institutional Research Ethics Board. Samples were collected fresh from the patients at the time of surgical resection and immediately snap frozen in liquid nitrogen and stored at -80°C . Healthy meninges were collected from patients who were underwent neurosurgery for non-oncological disease.

Clinical data was collected as per pre-established consensus definitions as indicated above. Briefly, for each case, hematoxylin and eosin (H&E) slides were reviewed to confirm the diagnosis of meningioma, grade tumors according to the current 2016 WHO criteria, and subtype tumors according to recognized histopathological classifications, where appropriate, by two experienced neuropathologists independently. Given the tendency for local aggressiveness in a subset of meningiomas, tumor recurrence and time to recurrence were the primary outcomes of interest in this study. Recurrence was defined as tumor growth following gross total resection or tumor progression following subtotal resection that resulted in a change in management and the time to recurrence was determined by calculating the duration from the date of surgery to first postoperative imaging documenting tumor recurrence. The extent of resection (Simpson grade) was extracted from the surgeon's operative report and checked with postoperative magnetic resonance imaging (MRI). Additional clinical information including but not limited to sex, age at surgery, previous treatment, post-operative treatment, and tumor location were annotated for each sample.

DNA and RNA processing

DNA and RNA were extracted from adjacent but regionally distinct tissue for each patient. DNA was extracted from tumor and matched normal tissue (whole blood) as well as healthy meninges samples using the DNeasy Blood and Tissue Kit (Qiagen, USA) and quantified using the Nanodrop 1000 instrument (Thermo Scientific, USA). Total RNA was isolated from tumor samples using the RNeasy Mini Kit (Qiagen, USA) and quantified using the PicoGreen assay. RNA integrity was assessed using the Agilent 2100 Bioanalyzer (RNA; Agilent, USA) and samples with RNA Integrity Number (RIN) > 7 were selected for further sequencing.

Genome-wide DNA methylation

Illumina Infinium MethylationEPIC BeadChip array (Illumina, San Diego, USA) was used to obtain genome-wide DNA methylation profiles on 250–500ng of bisulfite-treated DNA (EZ DNA Methylation Kit, Zymo, California, USA) per tumor and healthy meninges samples. Raw methylation files (*.idat) were imported, processed and normalized (ssNoob) using minfi¹⁸ (v1.34). Probes that failed to hybridize (detection p value > 0.01) in one or more samples were removed from downstream analyses. Probes that overlapped with known single nucleotide polymorphisms, cross-reactive probes, and probes that localized on X and Y chromosomes were also removed for all unsupervised analyses. Differentially methylated probes were identified using limma¹⁹ based modelling approach. When comparing meningiomas to healthy meninges, CpG sites were considered differentially methylated if the absolute mean differences in β -value > 0.35 and adjusted p-value (FDR-corrected) < 0.05. When comparing each molecular group to healthy meninges, this threshold was adjusted to absolute mean differences β -value > 0.1 and adjusted p-value (FDR-corrected) < 0.05. Probe annotation was performed using the UCSC Genome Browse (hg38 assembly).

Whole exome sequencing

Exome libraries were prepared using 100ng DNA of tumor tissue or matched normal DNA. Exome capture was performed using Agilent SureSelect Human Exome Library Preparation V5 or V6 COSIMC + kits and sequenced (pair-ended) on a HiSeq 2500 platform to a median of 191X. Raw sequencing data (fastq files) were aligned to the hg19 reference genome using BWA-MEM v0.7.12²⁰ with default parameters. PCR duplicate marking, indel realignment and base quality score recalibration were performed using Picard v1.72 and GATK v3.6.0²¹. Data quality assessment was performed using CalculateHSMetrics from Picard. Somatic mutations were identified using Mutect V1.1.7²² and Strelka v1.0.13²³ for tumors with matched peripheral blood controls and Mutect2 V1.1 for tumors without matched peripheral blood controls. All mutations in genes that are recognized drivers in meningiomas (i.e., NF2, SMARCB1, TRAF7, AKT1, KLF4, SMO, POLR2A, DMD) were retained for statistical analyses. For the discovery of novel, functionally-relevant genes, germline variants with GnomAD²⁴ population frequency >0.01% were removed to retain putative somatic mutations. Variants with allele frequency of > 10% and a TGL frequency database of variants of < 1% were retained to filter out initial passenger events. Genes with at least two somatic protein-altering mutations were selected, and the statistical basis for the filtered mutations were checked using MutSigCV²⁵ for the overall cohort.

We used a threshold of $FDR < 0.1$ to consider variants as driver events, as described by the MutSigCV developers²⁵. The functional effects of variants were subsequently annotated using Variant Effect Predictor v.92.0²⁶, OncoKB Precision Oncology Knowledge Base²⁷, [CancerHotspots.org](https://cancerhotspots.org)²⁸ and dbNSFP database²⁹. Statistically significant variants that were predicted to be actionable/driver alterations or whose effects were predicted to be pathogenic/likely pathogenic are reported and shown in the Oncoprint in this manuscript. Tumor mutation burden was calculated as the fraction of total number of protein-altering (nonsynonymous) somatic mutations across the callable exome space (in Mb).

Gene-level copy number profiling

To assess allele specific copy number profiles, we used Sequenza v2.1.2¹⁹ for tumor-normal pairs and CNVkit v0.9.6³⁰ for unmatched tumor samples using a pooled reference set of 60 peripheral blood samples from individuals unrelated to the study. We used conventional thresholds set by cBioportal³¹ to classify chromosomal gains and deletions ($\log_2\text{ratio} > 0.7$ as a high-level gain and < -0.7 as a deep deletion). The degree of genomic disruption per sample was computed as the fraction of the genome that was affected by copy number gains or losses.

RNA sequencing

mRNA libraries were generated using NEB Ultra II directional mRNA library prep kit according to manufacturer's protocol. Libraries were sequenced on the Illumina HiSeq 2500 high output flow cell ($2 \times 126\text{bp}$), sequenced with 3 samples per lane to obtain approximately 70 million reads per sample. Raw sequencing data (fastq files) were processed and aligned to human reference genome (GRCh38) using STAR (v2.6.0a)³². Duplicate reads were removed, and reads were sorted using SamTools (v1.3³³). Raw gene expression counts were computed for each sample using featureCounts in the package Rsubread (v1.5.0³⁴) and subsequently normalized by counts-per-million (CPM) and subjected to TMM (trimmed mean of M) normalization using edgeR (v3.22.3)³⁵. TMM removes genes with low counts by cpm-cutoff to filter out noise. The values for cpm-cutoff were determined empirically by identifying the minimum value required to achieve the best normalization across samples. Using only protein-coding genes, the best cpm-cutoff was determined to be 1.

Mutual information analysis

The mutual information (MI) metric¹³ was computed for each gene using all pairwise combinations of molecular data in our study (DNA methylation, copy number alterations, mRNA abundance, protein abundance). MI measures the amount of information that is known about a gene by one datatype, when the paired datatype is already known. Conceptually, MI is related to classic correlations (such as Spearman or Pearson correlations), however, statistical assumptions regarding linearity and ordering are not absolute making this approach appropriate for modeling of complex relationships such as in cancer genomics. MI values of zero indicate completely independent variables, such that knowledge of one variable has no bearing on the knowledge of the other. For each pairwise comparison, data were discretized into 21 bins for each gene, and the mutual information (MI) between two datatypes was defined as:

$$MI_{xy} = H_x + H_y - H_{xy}$$

where H_x and H_y , the marginal entropies of datatypes x and y and H_{xy} is the joint entropy calculated using the R package Entropy (V.1.2.1). MI was normalized over the mean entropy of the two input vectors. To assess statistical significance of normalized MI values, permutation testing was performed. Gene-level data were permuted 100,000 times to generate a null MI distribution and p values were calculated as the proportion of null MI values greater than or equal to the true observed MI. P-values were FDR-adjusted and the significance threshold was set at an FDR of 5%. Consensus clustering³⁶ was performed on those genes for which MI was significant for at least one datatype pair, after subsetting for genes with data available for all four datatypes. The divisive analysis clustering (diana) algorithm was applied to z-scored normalized MI values, using a maxK of 10 with 1000 resampling repetitions. For methylation data, the Pearson correlation between gene-level RNA abundances and corresponding probe β values was calculated, and the probe with the greatest negative correlation was selected. For genes with annotated probes but without corresponding RNA abundance measures, the probe with the highest variance in β across samples was selected. This was done to achieve a 1:1 gene:probe relationship.

Single platform clustering analyses

In order to identify the optimal number of clusters using mRNA data, gene-level somatic copy data, and DNA methylation data, we performed Consensus Clustering using the ConsensusClusterPlus³⁶ R package for each individual datatype separately. Consensus Clustering was performed using the top 5,000 most variably expressed genes, 1000 most variably altered genes, and 10,000 most variably methylated CpG sites as determined by median absolute deviation of logCPM, log₂CNVratios and β -values across all samples for RNA seq, gene-level copy number, and DNA methylation data, respectively. Clustering was performed using Pearson correlation for the distance metric and Ward linkage algorithm with 1000 resampling repetitions (epsilon=0.8). For each platform, we computed the average silhouette width as well as plots of the Cumulative Distribution Function (CDF) of the consensus matrix for each k subgroups to identify the optimal k where CDF reaches an approximate maximum. For gene-level copy number and gene expression we determined the optimal $k=6$. For DNA methylation data, both $k=5$ and $k=6$ provided similar results. Given previous reports of $k=6$ methylation subgroups, we selected $k=6$ as the optimal number of methylation-based clusters. Samples were then projected into a two-dimension space using t-distributed Stochastic Neighbor Embedding (tSNE) for cluster assignment and visualization for each individual platform separately. Divergence from expected recurrence-free survival patterns in our samples using a previously established methylation-based cluster classification³ led us to use data-driven methylation cluster groupings for our analyses in this paper. Adjusted Rand Indices (ARI) were calculated on cluster assignments for each pairwise combination of datatypes to determine the degree of cluster overlap.

Cluster of Cluster Assignments

To comprehensively integrate mRNA, copy number, and DNA methylation data, we employed the Cluster-of Cluster Assignments (COCA) algorithm that has been used by the TCGA to identify molecular subtypes of systemic cancers^{9–12}. Cluster assignments from unsupervised tSNE-based individual platform clustering were first binarized into indicator variables that were combined to construct a matrix-of-clusters (columns are binarized cluster memberships and rows are samples). This second order matrix was then subjected to an additional round of consensus clustering to examine the relationship of samples across molecular features. The optimal number of subgroups was selected by computing and maximizing the average silhouette width from $k=2$ to $k=10$. To examine the relative importance of each datatypes, COCA was repeated with all combinations of two datatypes at a time. Cluster assignments by integration of three versus two datatypes were compared for overlap by computing Adjusted Rand Indices (ARI).

Cancer cell fraction (CCF) estimation

The cancer cell fraction of variant i (CCF_i) was calculated as follows:

$$CCF_i = \frac{u_i}{m_i}$$

where u_i is a function of the variant allele fraction of variant i (f_i), sample purity (ρ), the local copy number of the tumor cells at site i ($n_{total,t,i}$) and the local copy number of the normal cells at site i ($n_{total,n,i}$, assumed to be 2)³⁷.

$$u_i = f_i \frac{1}{\rho} [\rho n_{total,t,i} + (1 - \rho) n_{total,n,i}]$$

The variant allele fraction of variant i (f_i) was directly calculated using the number of reference reads for locus i ($r_{ref,i}$) and the number of alternate reads for locus i ($r_{mut,i}$).

$$f_i = \frac{r_{mut,i}}{r_{mut,i} + r_{ref,i}}$$

For each sample, we estimated sample purity (ρ) as previously described using DNA methylation data³⁸. The local copy number of the tumor cells at site i ($n_{total,t,i}$) was transformed from the segment mean at site i (s_i).

$$n_{total,t,i} = 2^{1 + s_i}$$

The mutation multiplicity of variant i (m_i) was determined using the following equation:

$$m_i = \begin{cases} |u_i| & u_i \geq 1 \\ 1 & u_i < 1 \end{cases}$$

Finally, if the CCF_i was greater than 0.80, then, variant i was considered clonal.

Differential gene expression analysis

Differential gene expression analysis was computed using gene-wise negative binomial generalized linear models with quasi-likelihood tests (F-test, edgeR³⁵ v3.22.3). Genes were ranked by combining direction of fold changes and computed p-values using the following formula: $\text{sign}(\log\text{FC}) \times -\log_{10}(\text{p-value})$, where $\text{sign}(\log\text{FC})$ determines the direction of the change (upregulated as positive and downregulated as negative) and $-\log_{10}(\text{p-value})$ determine the magnitude of ranking. Gene-Set Enrichment Analysis (GSEA, v3.0) was performed as previously described using ranked scores as input to determine if differentially expressed genes belong to common biological pathways³⁹.

Pathway analysis and network maps

Pathway analyses and network maps were generated as previously described³⁹. Pathways were defined by the gene set file Human_GOBP_AllPathways_no_GO_ia_June_20_2019_symbol.gmt that is maintained and updated regularly by the Bader lab (http://download.baderlab.org/EM_Genesets/). GeneSet size was limited to range between 10 – 200 and 2000 permutations were carried out. The results of the pathway analysis were visualized using the EnrichmentMap App (v1.2.0) in Cytoscape (v3.7.2). Network maps were generated for nodes with FDR q-value < 0.01, p-value < 0.0001, and nodes sharing gene overlaps with Jaccard Coefficient > 0.25 were connected by a green line (edge). Clusters of related pathways were identified and annotated using a Cytoscape app that employs a Markov Cluster Algorithm that connects pathways by shared keywords in the description of each pathway (AutoAnnotate, v1.2). The resulting groups of pathways are designated as the major pathways in a circle.

FDA drug mapping

In order to discover realistic and novel therapeutic agents, we examined whether FDA approved drugs could be repurposed for the treatment of meningioma by examining for the presence of FDA approved drug targets in our network analyses. Drugs were selected by the number of target genes in the leading edge of significant GSEA pathways for indicated comparison, then each drug was ranked by the number of genes plus pathways targeted. Finally, the number of significant genes targeted were divided by the total number of target genes of the drug to assess the specificity. This scoring system selected the drugs targeting the greatest number of driving genes in significant biological pathways with high specificity. The resulting list of drugs were grouped by common targets to produce a higher-level summary of the class of drugs with the highest possibility of effective treatment. Individual drugs were visualized on pathway maps using Post-Analysis function in the Enrichment Map plugin of Cytoscape app.

Gene fusion identification

Interchromosomal and intrachromosomal gene fusion events were detected using FusionCatcher v1.1.0, with default parameters. FusionCatcher aligns reads to the human reference genome (GRCh38) using Bowtie⁴⁰ (v1.2), Bowtie2⁴¹ (v2.3), BLAT⁴² (v0.35) and STAR BLAT³² (v2.7). Adjacent and read-through fusions were filtered out from analyses and fusions with Counts_of_common_mapping_reads = 0 were selected to reduce false positive detection of genes with similar sequence homology. A stringent threshold for conservative estimation of fusion events (unique spanning reads ≥ 25) was used to assess interchromosomal and intrachromosomal fusion events.

Generalization cohort

Large ($n > 50$), multi-omic meningioma datasets in the literature with matched individual patient outcome data were not available for use as independent validation. Therefore, to confirm the generalizability of the association with integrative MGs and their association with outcomes, we assembled an independent cohort of 80 meningioma patient samples with longitudinal outcome data and generated mRNA-sequencing data. Assignment of MG for each new sample was performed by a single-sample Gene Set Enrichment Analysis (ssGSEA) using the top 50 highly expressed genes for each group in the initial discovery cohort. Cluster assignment was determined by maximal scores from ssGSEA analysis and checked by unsupervised hierarchical clustering of ssGSEA scores. Kaplan–Meier estimates of survival with log-rank tests for association were performed to test the association of MGs in the new independent cohort with outcome. The association of MGs with outcomes were compared to WHO grade by generation of Brier prediction curves and computation of Brier scores.

The discriminative capacity of gene expression profiles to distinguish MG groups overall was quantitated using true gene expression classifiers (generalized linear model, default alpha and lambda parameters) for each MG in the discovery cohort. To do this, we randomly split out cohort into training and test sets, with 90% of the data in the training and the remaining 10% of the data in test set. Expression classifiers for each MG group were trained using the top 50 highly expressed genes for each MG, and the performance for each classifier was tested using held-out samples in test cohort by computing the area under the receiver operative characteristic curve (AUC). This process was repeated for a total of 50 iterations of training and testing

Epigenetic mitotic clock analyses

We used previously described mitotic clocks (epiTOC¹⁶, epiTOC2¹⁵, and solo-WCGW¹⁴) that are based on DNA methylation to examine regions of the genome that are either fully methylated or unmethylated in multiple fetal tissues but gain or lose methylation as a function of mitotic age. The epiTOC model calculates a weighted average methylation over 354 CpGs on the 850K array at gene promoters marked by the PRC2 complex that are constitutively unmethylated in fetal tissue and increase in methylation with age and cell division. The epiTOC2 model estimates the mitotic age (adjusted for chronological age of patient) using a weighted subset of 151 CpGs from the epiTOC model that are most likely to change in DNA methylation levels with age. The solo-WCGWs are a set of CpGs at

the WCGW motif without flanking CpGs that are hypomethylated in fetal tissues and gain methylation with age and cell division. A total of 6214 solo-WCGWs that were originally described are found on the EPIC array. Of note, 648 of these are uniformly hypomethylated across multiple fetal tissue types, as previously described, and therefore a weighted average of these 648 CpG sites was used to derive the “HypoClock” score.

Transcription factor analyses

We identified master transcription factors for each MG as previously described using ElmerV2¹⁷. First, differentially methylated distal CpGs at non-promoter (i.e. probes further than 2kb from the transcription start site) sites were identified between each MG and every other MG independently as well as all other MGs as a group. Putative target genes were identified for each differentially methylated CpG by computing the correlation between methylation of the probe and the expression of the closest 10 upstream and 10 downstream genes. Motif occurrences were identified using HOMER within 250bp region for significantly hypomethylated probes with putative gene targets and enrichment for motifs are calculated by computing the Odds Ratio (and 95% CI) that each probe in a probe set contains motif occurrences in comparison to a background of all distal probes on the 850K array. Transcription factors were considered enriched if the lower bound of the 95% CI was greater than 1.1. Finally, the mean methylation of all probes in probe-gene pairs that contained a given motif instance within 250bp were compared to the average expression of a set of 1639 transcription factors^{43,44}. These were then ranked by degree of anticorrelation using $-\log_{10}(\text{FDR})$ in order to identify master regulator transcription factors by transcription factor subfamily.

Shotgun proteomics

Approximately 1–2mg of fresh frozen meningioma tumors were pulverized using a Covaris cryoPREP Pulverizer. Pulverized tissue was then solubilized in 300 μL of 50% (v/v) 2,2,2-Trifluoroethanol in phosphate buffered saline (pH 7.4) with a 5min incubation at 95°C, repeated probe sonication, freeze-thaw cycling, followed by a two-hour heated incubation at 60°C. 100 μg of protein lysate was denatured with 5mM dithiothreitol for 30min at 60°C and reduced disulfite-bonds were subsequently alkylated with 25mM iodoacetamide for 30min at room temperature in the dark. Proteins were digested into peptides with 4 μg of trypsin at 37°C overnight. Peptides were then desalted and purified using C18-based solid phase capture. Eluted peptides were lyophilized and solubilized in mass spectrometry-grade water with 0.1% methanoic acid and peptide concentration was quantified using a NanoDrop Lite spectrophotometer (at 280nm). For each sample, an Easy1000 nanoLC was used to load 2 μg of peptides onto a 2cm trap column (Thermo Scientific). The peptides were separated along a four-hour gradient using a 50cm EasySpray analytical column coupled by electrospray ionization to an Orbitrap Fusion (Thermo Scientific) tribrid mass spectrometer. Peptides were detected using a Top25 data-dependent acquisition method. The acquired data was searched using Maxquant (v1.6.2.3⁴⁵) against a UniProt complete human protein sequence database (v2019_04) with an FDR of 1% for peptide spectral matches. Two missed cleavages were permitted along with the fixed carbamidomethyl modification of cysteines, the variable oxidation of methionine and variable acetylation of the protein N-terminus. Relative label-free protein quantitation was calculated using MS¹-level peak integration

along with the matching-between-runs feature enabling a 2min retention time matching window. Proteins identified with a minimum of two peptides were carried forward for further analysis. Protein-groups with $\text{Log}_2\text{FC} > 2$ i.e. 4-fold higher expression or more, and $\text{FDR} < 0.05$ were considered specific for each group.

Validation of proteomic findings by immunohistochemical analyses

To validate the enrichment of group-specific proteins identified by proteomic data, we performed immunohistochemical analyses for S100B, SCGN, ACADL, MCM2 in a cohort of 44 tumors with known MG status. Experimentation and analyses were performed blinded to MG status. Briefly, consecutive 5-micron formalin fixed, paraffin sections were rehydrated and heat-mediated antigen retrieval using sodium citrate buffer (pH 6) was performed. Slides were washed in 3% H₂O₂ in methanol and blocked in 5% BSA in PBST for 1 hour at room temperature followed by overnight incubation at 4°C with anti-S100B (ThermoFisher, #701340, dilution 1:100), anti-SCGN (Sigma, HPA006641, dilution 1:500) anti-ACADL (Sigma, HPA011990–100UL, dilution 1:200) or anti-MCM2 (Cell Signalling, #12079S, dilution 1:200). The expression signals were developed using DAB Peroxidase Kit and the slides were counterstained with hematoxylin, dehydrated, and coverslipped. Whole slide images were digitized and obtained using virtual microscopy. Tumor tissue was annotated in each whole slide by an experienced and blinded neuropathologist and subsequently subjected to unbiased quantitative digital pathological assessment using the Multiplex IHC module on HALO software (Indica Labs, Albuquerque, NM, USA).

Droplet-based single nuclear RNA-sequencing (snRNA-seq)

Ten Frozen archived tumor specimens and two frozen archived healthy meninges were minced with sterile scalpel and homogenized using a dounce tissue grinder (size A and B, Sigma Aldrich) in ice cold lysis buffer (0.32 M sucrose, 5mM CaCl₂, 3 mM MgAc₂, 20 mM Tris-HCl, 0.1 mM EDTA, 40U/ml RNase inhibitor and 0.1% Triton X-100 in DEPC-treated water). Homogenized tissue was centrifuged at 500g for 10minutes at 4°C, washed in two rounds using ice cold wash buffer (1x PBS, 12mM EGTA pH 8.0 and 0.2 U/μl RNase Inhibitor) and the nucleus pellet was subsequently resuspended in resuspension buffer (1x PBS, 0.04% BSA) prior to filtration using 40μm Flowmi cell strainer (Sigma Aldrich). Isolated nuclei were stained with DAPI and fluorescence-sorted (BD Influx BRV, Becton Dickinson Biosciences) to retain healthy nuclei. DAPI+ nuclei were washed and resuspended in resuspension buffer. Nuclei were counted and approximately 6000–8000 nuclei were loaded onto a 10x Chromium controller using the Chromium Single Cell 3' Library & Gel Bead v3 (10x Genomics) for each sample. Single nuclei were partitioned into barcoded Gel Beads in Emulsion (GEMs) in the Chromium instrument, followed by cell lysis and reverse transcription of RNA in the droplets. Breaking of the emulsion was followed by cDNA amplification and library construction as per manufacturer's recommendations. Samples were sequenced Illumina NovaSeq (10x specific protocol) with a median target sequencing depth of 60,000 reads/nuclei.

snRNA-seq raw data processing, filtering and validation of cells to patients

Raw sequencing data (bcl files) were converted to demultiplexed fastq files (Illumina bcl2fastq, v2.19.1) and aligned to the human genome reference sequence (hg38). Expression

matrix of unique molecular identified (UMI) counts per gene per nuclei was obtained using CellRanger (10x Genomics). As the first step for validating cells to patients, we looked to confirm that cells had data that covered known single nucleotide polymorphism (SNP) regions. To do this, we quantified the number of unique molecular identifiers (UMI) mapping to a panel of 7.4 million SNPs identified through the 1,000 Genomes Project⁴⁶ with minor allele frequency > 5% using cellsnp-lite. Two of our samples had highly sparse coverage of known SNP regions and were not reliably genotypable and therefore removed from further analyses.

To validate the assignment of cells to patients for samples that had potential overlap in processing, we compared SNPs derived from single cell RNA seq data to SNPs derived from bulk RNA-seq data using *demuxlet*⁴⁷. Demuxlet was developed to deconvolute sample identity when multiple samples are pooled by barcoded single cell sequencing. Variant call format (VCF) files from bulk RNA-seq data were generated and compared to variants identified in single cell data by *demuxlet*. Only cells with genotypes that aligned to the expected sample were retained for further analyses. Potential doublets were identified using *scDbtFinder* (v3.13) and removed.

From all remaining cells, we quantified two quality measures for each cell: the number of UMIs detected, and fraction of mitochondrial transcripts. Low-quality cells where >1.5% of transcripts derived from the mitochondria and cells with low complexity libraries in which less than 1000 UMIs were detected were removed. Following data filtering a total of 54,393 high-quality single nuclei that were genotyped to 10 samples were retained for analyses.

snRNA-seq clustering of all cells

Library size normalization was performed as previously described using *scraper* where hierarchical clustering of cells using Spearman distances subset cells into more groups, and then scaling factors per cell were determined by randomly pooling cells, computing summed library sizes, and comparing to average library size across all cells in each group.^{48,49} Normalized UMI counts were used for clustering by optimizing a shared nearest neighbour (SNN) modularity function with Seurat⁵⁰. First, principal component analysis was performed using highly variable genes (FDR < 0.001) identified by *scraper*. The number of significant Principal Components (PC, 10) was determined based on the inflection point of a 'scree' plot. Next, a shared nearest neighbor graph was built from distances computed in first 10 PC space and clusters were identified by optimizing the modularity function within this space with a resolution set to 0.1. Gene expression and clustering results were visualized using t-distributed Stochastic Neighbor Embedding (t-SNE) of the selected principal components.

Cell type classification

Cells were assigned to different cell types based on consensus by:

1. Similarity of expression profiles: As neoplastic and stromal/immune compartments are expected to have different expression profiles, we first correlated (Pearson) the expression profile of each cell to every other cell. Unsupervised hierarchical Pearson clustering with Ward linkages on the matrix

of correlation values was performed and two major clusters (putative neoplastic and non-neoplastic) of cells were identified

2. Copy number profiles: We used inferCNV(v1.1.1)⁵¹ to infer copy number alterations of neoplastic and non-neoplastic cells with snRNA-seq data. Cells from healthy meninges were used as the reference set. Genes were ordered from the human GRCh38 assembly, and a heatmap illustrating relative expression intensities of neoplastic nuclei to reference population across the genome was generated for visualization. Almost all neoplastic clusters harbored loss of chromosome 22q that was not observed in non-neoplastic cells that were generally devoid of significant CNA. We further computed a general metric of aneuploidy using inferred CNA data by first scaling CNA to the range of -1 to 1 , and then summing the absolute copy number ratios for all genes. The degree of aneuploidy was later used to compare cells of high versus low potency.
3. Expression of canonical markers: Significantly differentially expressed genes were identified for each cluster using *FindAllMarkers* in Seurat and these were inspected for canonical immune and stromal cell markers. Enrichment of these markers across clusters was visualized by bubble plots and was indicative of cell type annotation. Predictions regarding cell cycle phases were made for neoplastic cells on the basis of the expression of a core set of genes, as previously described⁵².

Correlation of CNA inferred from snRNA-seq data and bulk WES data

To correlate CNA data from snRNA-seq and bulk WES data, inferred CNA ratios from snRNA-seq was scaled to values between -1 and 1 such that the two datasets were similarly scaled. Arm-level copy number ratios were then computed from snRNA-seq and bulk CNA data independently, as follows:

$$CN_s = \frac{\sum(CN_i \times L_i)}{\sum L_i}$$

where CN_i is the copy number ratio of the i^{th} gene in segment s and L_i is the length of the i^{th} gene. Pearson and Spearman correlations were then computed on arm-level CNA ratios from both datatypes.

snRNA-seq clustering of individual samples

To examine heterogeneity within tumors, we clustered cells from each patient independently using two independent approaches (Seurat and DBSCAN). Clustering by Seurat⁵⁰ was performed as described above, with resolution set to 0.05 to account for the smaller number of cells with single sample analyses.

DBSCAN identifies clusters by identifying dense regions in space, ensuring that the neighbourhood of a radius (epsilon) has to contain minimum number of neighbours (minPts). DBSCAN identifies outliers of cells that do not belong to any clusters (considered

noise). To cluster cells by DBSCAN we first normalized raw expression levels for each sample as follows:

$$E_i = \frac{\log_2(1 + CPM_i)}{10}$$

where CPM_i for genes i to n was computed as $10^6 \times UMI_i / \Sigma(UMI_{1...n})$. These values were then centered to the average expression of the gene across all cells in the sample to define relative expression of each gene in each cell. Using this data, each sample was subjected to dimensionality reduction by tSNE (with a perplexity of 30) followed by density clustering using DBSCAN (parameters $\epsilon=1.8$ and $\text{minPts}=5$). Cells that did not meet these parameters were considered unclassifiable and colored gray in the tSNEs.

Statistical evaluation of between and within patient variation

We used a one-way ANOVA test on the top 10 principal components for all neoplastic cells to compare between patient variability and within patient variability as previously described⁵³. The F-statistic by ANOVA divides the variability observed in the dataset to between patient components and within patient components. F-statistics > 1 indicating that the between-patient variation is greater than the within-patient variation.

Statistical evaluation of two cell features

To examine whether two features of a cell were associated, we used mixed-effects logistic regression models that are able to account for cell to patient dependencies, as previously described⁵⁴. We specifically used these models to test for the enrichment of immune cells in MG1, the enrichment of cycling cells in MG3 and MG4.

Non-negative matrix factorization to identify intrinsic gene expression programs

To identify intrinsic expression program, we applied non-negative matrix factorization (NMF) to relative expression levels used for DBSCAN analyses after transforming all negative values to zeros, as previously described⁵⁴⁻⁵⁶. Factors k ranged from six to nine and genes were ranked by NMF scores for each expression programs identified. A total of 39 expression programs were identified across eight tumor samples. We then performed hierarchical clustering of programs using the extent of shared genes as a distance metric (using the top 50 genes in each program) to identify meta-signatures that were recurrent across samples. We calculated the Pearson correlation coefficient between NMF scores and the fraction of mitochondrial genes to assess for the relationship of each program with technical confounders. One cluster of programs (25–39) showed higher positive correlation with fraction of mitochondrial genes quantitated. This was confirmed by manual inspection of the genes that showed several mitochondrial and ribosomal genes that highly score in these programs. These programs were excluded from further analyses as they were favored to reflect technical artifacts. We then computed activation scores of each NMF program from all cells using AUCell³⁴(v1.8.0) and compared the distribution of activation scores across tumors.

Deconvolution of bulk RNA-seq data using snRNA-seq signatures

We used CIBERSORTx⁵⁷(v1.0) to deconvolute bulk mRNA-seq data from all samples in this study. We first used CIBERSORTX to generate a gene signature matrix for each single cell cluster from our single cell RNA seq data. Genes with weights greater than 400 were selected for each cluster and used in consensus k-means clustering with 5000 repeats to partition bulk RNA-seq data into four groups for comparison with bulk molecular classification.

We then generated a signature matrix for each cell type (macrophage, Tcell, endothelial cell, fibroblast, neoplastic) using CIBERSORTx, and then used this to determine cell type composition of each of our samples with bulk RNA sequencing data using single cell Correction S mode with 100 permutations.

Patient-derived cell lines

Fresh tumor specimens were obtained intraoperatively from five patients in whom informed consent for tissue banking was obtained previously. Cell suspensions were created and maintained as previously reported (PMID 26174772) on ThermoFisher BioLite 100 mm Tissue Culture dishes in DMEM/F12 (Life Technologies, #10565) supplemented with 1mM non-essential amino acids 1 mM NEAA (Life Technologies, #11140), 100 U/mL Anti-Anti (Life Technologies, #15240) and 10% fetal bovine serum (Life Technologies, #16141) in a humidified atmosphere with 5% CO₂. Once confluent, cells were passaged following trypsinization. DNA and RNA were extracted from an aliquot of each cell line. DNA was subjected to bisulfite conversion for DNA methylation profiling. To demonstrate that these cell lines are faithful models of meningiomas, we compared the genome-wide methylome profiles of cell lines to meningiomas from our cohort as well as a panel of published 2798 tumors from 40 brain tumor types⁵⁸. We found that all cell lines in this experiment clustered together with human meningioma tumors. As well, classification of our cell lines using a publicly available DNA methylation-based random-forest model (DKFZ [MolecularNeuropathology.org](https://www.molecularneuropathology.org) online classifier version 3.1.5) assigned all primary patient-derived cell lines into the meningioma methylation class with high calibrated scores from (0.97–0.99). To assign cell lines to MGs, we generated mRNA-seq data from cell lines and performed single-sample Gene Set Enrichment Analysis (ssGSEA) using the top 50 highly expressed genes for each MG from the cohort of tumors in our dataset. Cell lines were assigned to MG by maximal ssGSEA scores.

Cell viability assay

Meningioma cells (ranging from 1500–4500 cells based on the plating efficiency of each cell line) were plated in technical triplicates in Corning 96-well white-walled plates. Cells were treated with vorinostat (SAHA/MK0683, InvivoChem catalogue No. V0255; diluted to concentrations 100 nM, 500 nM, 1000 nM, 5000 nM) or 5-azacytidine (InvivoChem catalogue No. V0404; 10 nM, 50 nM, 100 nM, 500 nM, 1000 nM) for 10 days. A medium-only control was used for each replicate of each drug treatment, and a DMSO-control was used for vorinostat and 5-azacytidine treated cells. Three separate biologic replicates separated by at least one passage of each cell line were completed. Following the completion of treatment, CellTitre-Glo luminescent cell viability assay was performed on all

samples in accordance with the manufacturer's instructions (Promega, #G7570). Cells were incubated for 10 minutes with the CellTitre-Glo reagent and luminescence was measured using a 96-well plate reader (GloMax-96 microplate luminometer; Promega). Background luminescence was measured in blank wells with medium without cells and subtracted from experimental values automatically. Statistical analyses of intergroup differences between cell lines at each dose of each respective drug were performed using a two-way ANOVA followed by Tukey's test.

In vivo patient-derived xenograft

For intracranial xenograft experiments, 1×10^6 of MG4 patient-derived cells were injected into the subdural space of NSG mice. Mice were anesthetized and their cranium were fixed in a stereotaxic frame. An incision was made 3-mm lateral to the midline on the right side of the skull. The bregma was visualized and a burrhole was drilled using an automated 1.5 mm drillbit 3-mm lateral and 1-mm anterior to the bregma. Cells were injected at a depth of 1-mm to the skull surface using a 26-gauge needle and stereotactic Hamilton syringe in 5 μ L of media over 3 minutes. Following injection, the syringe was slowly removed over 2 minutes to limit reflux of cells. The incision was closed with 6-0 absorbable sutures and Vetbond tissue adhesive was applied on top. Mice were treated with either vorinostat (50 mg/kg 1:1 DMSO:PBS) or vehicle control (1:1 DMSO/PBS at equivalent volume) via intraperitoneal injection daily for 10 days, starting on post-implantation day 7. All mice were imaged at 3–5-days post xenograft implantation using a Bruker 7-Tesla preclinical MRI (STTARR imaging facility, Toronto, Ontario) to confirm intracranial implantation. Additional serial MRI scans were performed every 3–7 days based on availability of our imaging facility to document interval tumour growth. MRI volumetric analysis of tumours were performed by an individual blinded to treatment group using the Horos/OsiriX™ open source DICOM reader (GNU Lesser General Public License, Version 3 (LGPL-3.0)). Xenograft tumor were segmented on each MRI slice manually and then reconstructed automatically to obtain a volume measurement for each animal at each radiographic time point. Statistical analyses comparing the mean xenograft volume between the vorinostat-treated and control mice were performed at each time point using Mann-Whitney U-test, with statistical significance set at $p < 0.05$. Mice were sacrificed when they reached their physiologic or experimental endpoint in accordance with our animal care facility and the Canadian Council on Animal Care (CCAC) guidelines. Specifically, the endpoint was reached when mice lost $>20\%$ of their starting bodyweight, demonstrated significant lethargy and decreased activity, had visible cranial enlargement, or had tumour volumes exceeding 500 mm³ on MRI volumetric measurements. None of the animals in our study exceeded these endpoints without being mandatorily euthanized and no animal tumours achieved or exceeded the volumetric endpoint.

Survival Analyses

For comparison of survival between independent groups, Kaplan-Meier survival plots were generated using package survminer and log-rank tests were performed to test the null hypothesis of no differences between independent subgroups. Univariable Hazard Ratios (HRs) with 95% CI and p-values for clinical factors as well as MGs 1–4 were computed by fitting Cox Proportional Hazards Models. Multivariable survival analyses were performed

by fitting Cox Proportional Hazards Models that included for all factors significant on univariable analyses. Prediction error curves were generated to compare discriminative capacity of Cox Proportional Hazards Models by leave one out cross validation.

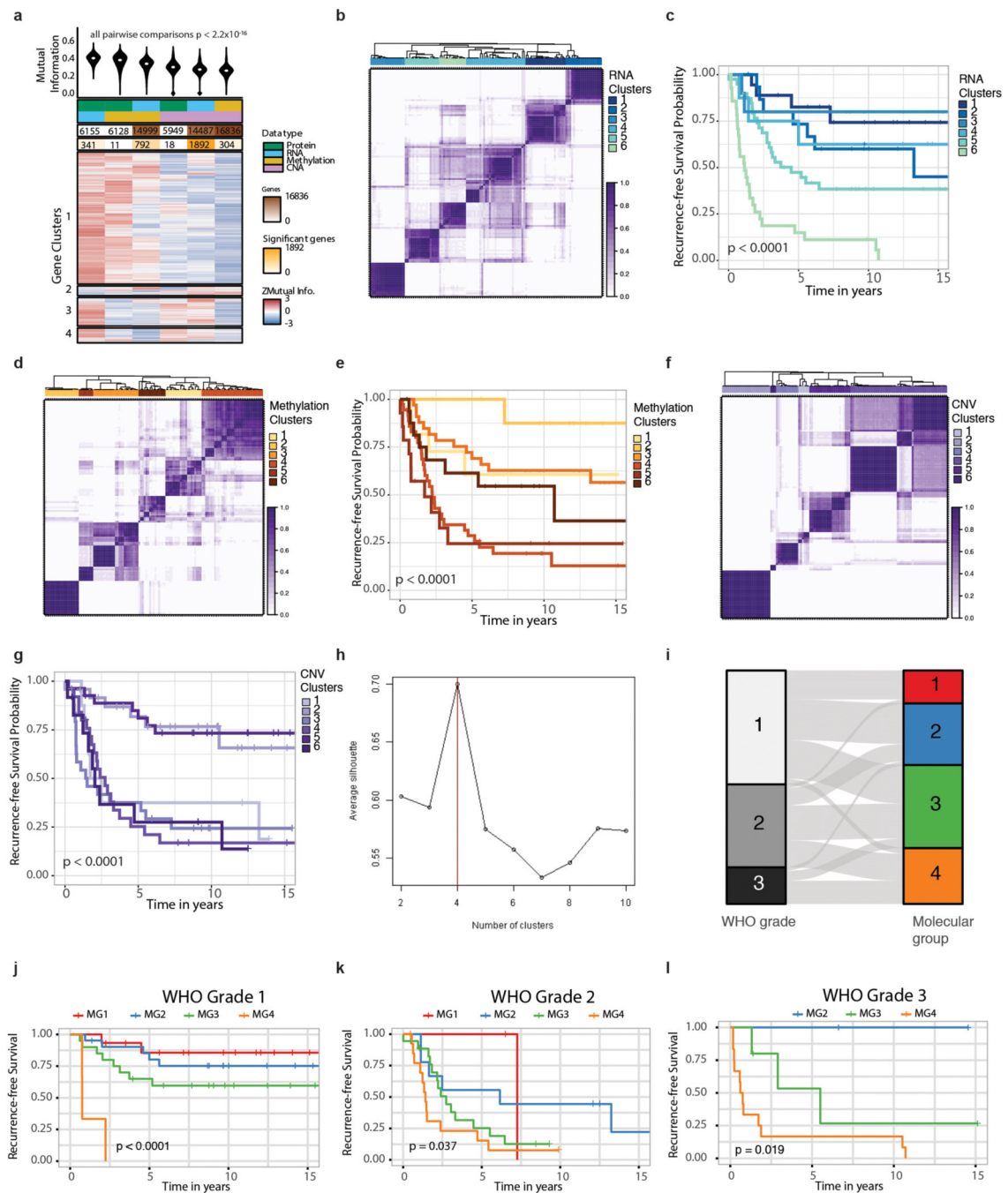
Data Availability:

Raw sequencing data for all datatypes have been deposited into public repositories. Proteomic data has been deposited to Mass Spectrometry Interactive Virtual Environment (MassIVE ID MSV000086901). DNA methylation idat files have been deposited to Gene Expression Omnibus (GEO, GSE180061). Whole exome-sequencing (fastq), bulk mRNA (fastq) and single nuclear RNA (fastq) datasets have been deposited to European Genome Archive under study ID EGAS00001004982 and dataset IDs EGAD00001007051, EGAD00001007494 and EGAS00001004982. The processed genomic data has been submitted to cBioportal at: https://www.cbioportal.org/study/summary?id=mng_utoronto_2021.

Code Availability:

Specific code will be made available upon request to gelareh.zadeh@uhn.ca

Extended Data



Extended Data Fig. 1. Individual datatype classification of meningiomas.

a, Violin plots showing the distribution of the normalized mutual information (MI) for each pairwise comparison of datatype. Median is shown as white dot. The number of total genes and number of genes with statistically significant (FDR < 5%) MI values are shown. Below this is a heatmap showing the consensus clustering of genes where

MI was significant for at least one datatype pair. Rows represent a gene for which data exists from all data types.

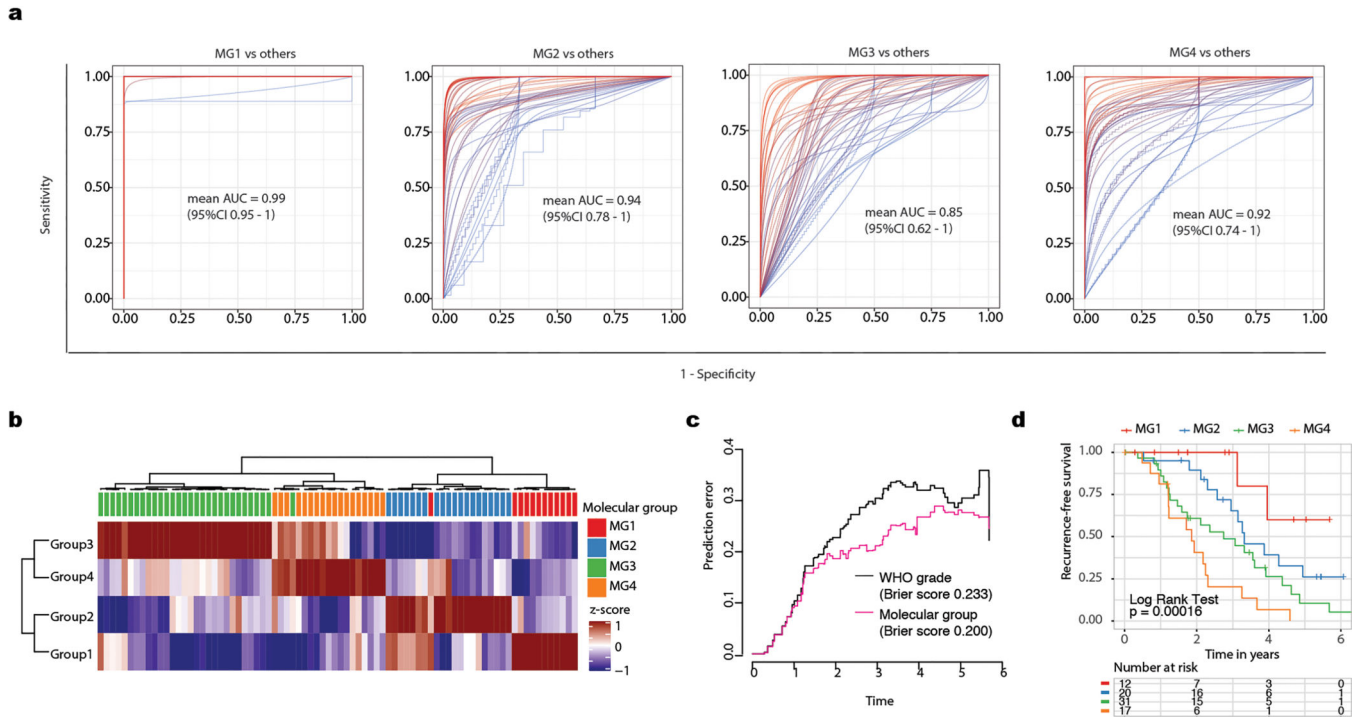
b,d,f, Unsupervised consensus hierarchical clustering of (b), 5,000 genes that show that highest median absolute deviation across expression values, (d), 10,000 CpG sites that show that highest median absolute deviation across β -values, (f), 1,000 genes that show that highest median absolute deviation across copy number ratios. Heatmap of consensus matrices with K=6 groups (b,d,f) are displayed. Overall, six groups were most stable across all platforms.

c,e,g, Kaplan Meier-plot displaying recurrence-free survival (RFS) distributions of unsupervised cluster assignments by (c) mRNA data, (e) DNA methylation data, (g) copy number data. The associations with outcomes are unique for the 6 cluster groups obtained on individual platform analyses.

h, Average silhouette widths for unsupervised consensus hierarchical clustering from K=2 to K=10. The silhouette score is a measure of stability of number of groups. Higher scores indicate greater stability and robustness. Average silhouette width is highest at K=4 subgroups.

i, Alluvial plot demonstrating associations between WHO grade and integrative molecular groups defined in this study.

j-l, Kaplan Meier-plot displaying recurrence-free survival (RFS) distributions of patients stratified and colored by molecular group assignments for WHO grade 1 tumors (j), WHO grade 2 tumors (k), and WHO grade 3 tumors (l).



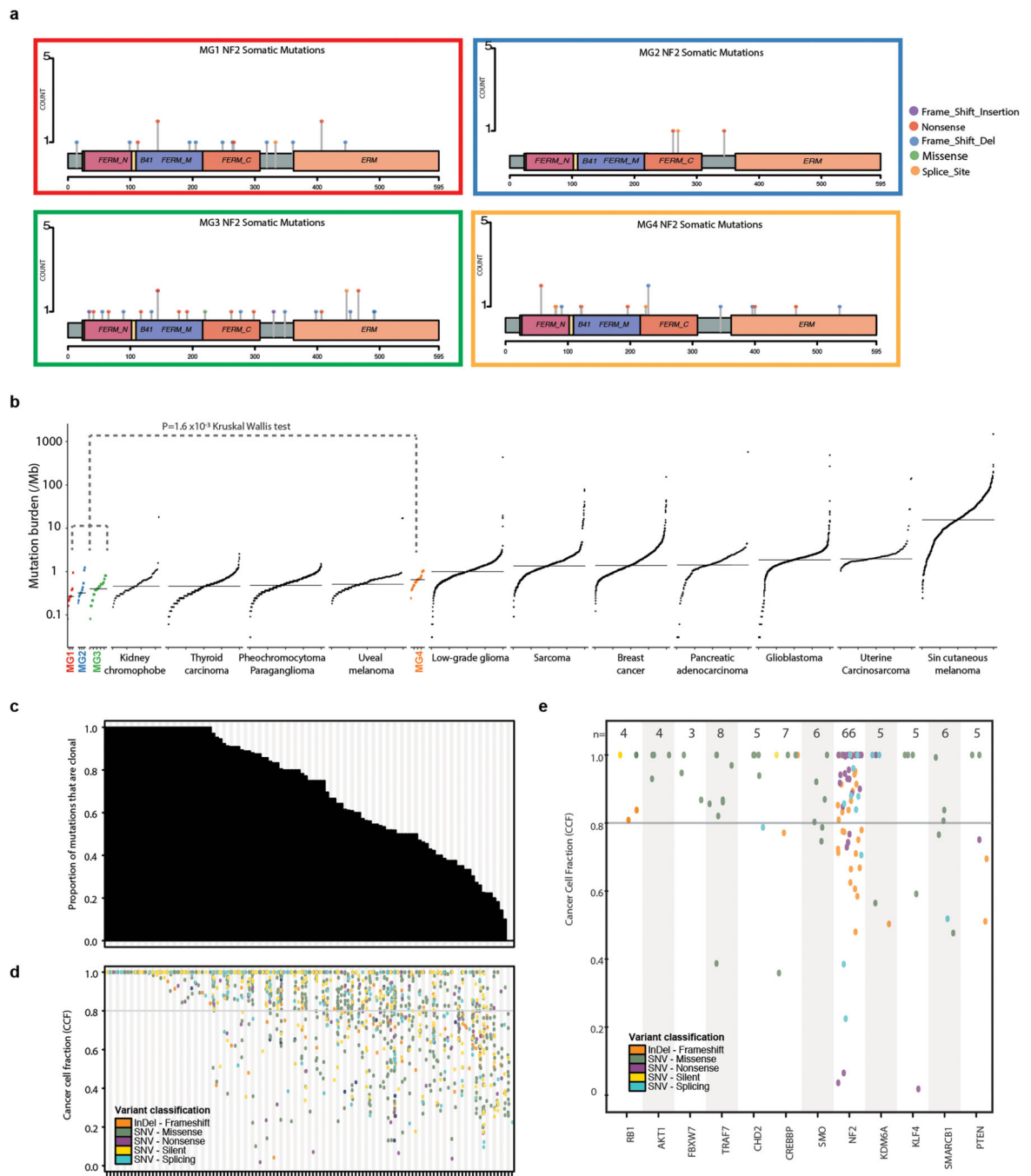
Extended Data Fig. 2. Generalizability of the association of molecular groups with outcome

a, Ensemble of Receiver Operating Characteristic (ROC) curves from 50 iterations of trained MG-versus-other models. Overlaid for each model is the mean Area Under the Curve (AUC) and its associated 95% confidence interval for samples in corresponding test sets.

b, Heatmap showing results of single-sample Gene-Set Enrichment Analysis (ssGSEA) using mRNA data in an independent cohort of 80 meningioma samples. Each sample in the validation set was assigned a score for Molecular Groups 1, 2, 3 and 4 using gene-expression based signatures from the discovery cohort. MG designation was determined by highest scores from ssGSEA assignments. Unsupervised hierarchical clustering using scores from MG assignments revealed four distinctive groups of tumors with 97% of samples having concordant assignment by maximal scores. Samples almost always showed high scores that were distinctive to only a single group, highlighting the robustness of classification in an independent cohort.

c, Brier prediction curve for recurrence-free survival comparing molecular group to WHO grade in the generalization cohort. The models tested were those developed on the discovery cohort. Prediction errors are consistently lowest using molecular groups in comparison to the validation cohort (integrated Brier score 0.179 vs 0.211).

d, Kaplan Meier-plot displaying recurrence-free survival (RFS) distribution of patients stratified and colored by molecular group assignments for generalization set. P value reported is a Log Rank Test. Distributions are highly similar to those obtained in discovery cohort.



Extended Data Fig. 3: Most mutations are clonal in meningioma

a, Lollipop plots showing the distribution of NF2 mutations by genomic regions within each molecular group.

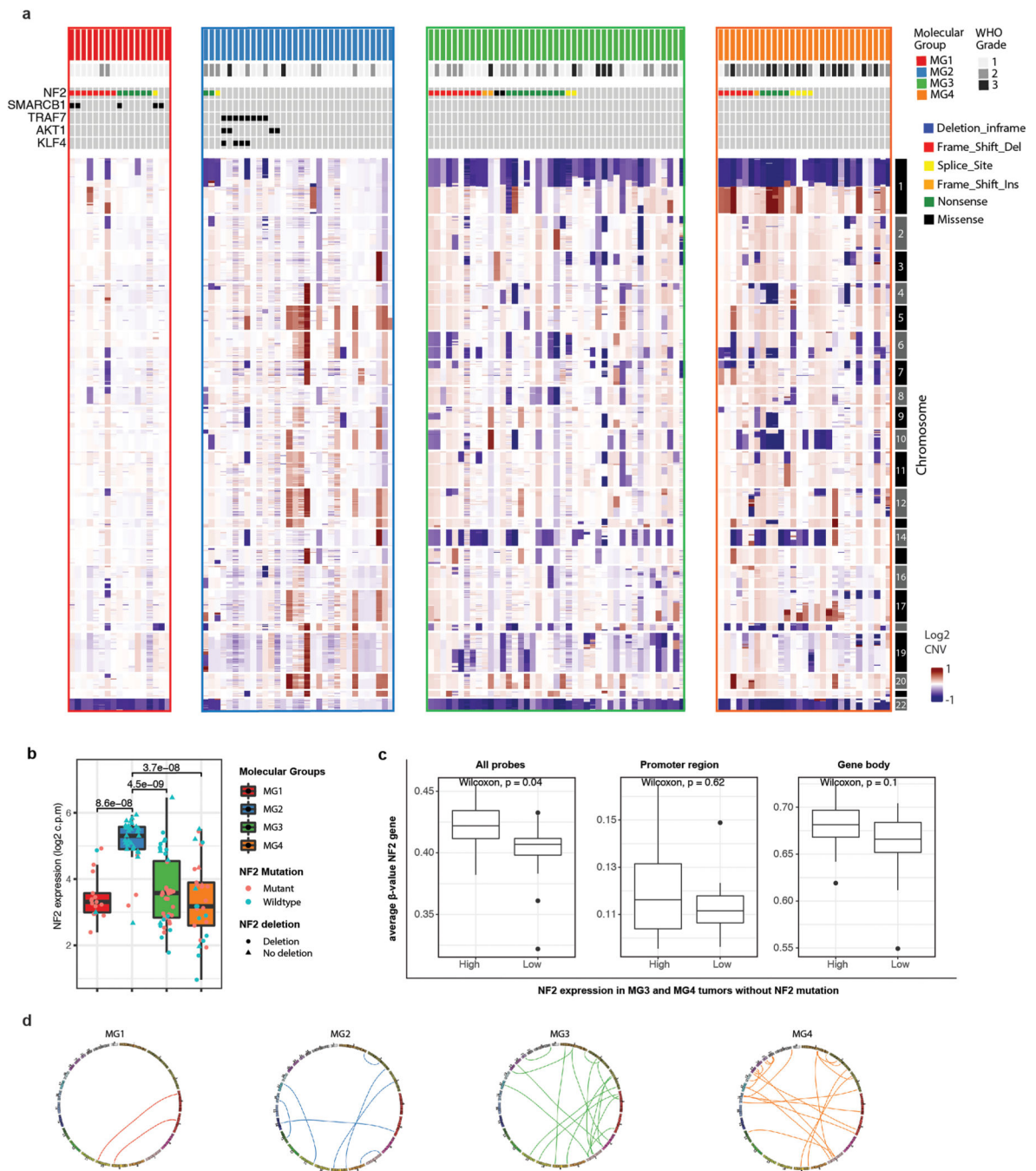
b, Mutational burden (nonsynonymous mutations per megabase) of meningiomas stratified by molecular groups in comparison to other TCGA solid cancers. Every dot represents a sample and horizontal lines are median number of mutations in each cancer type. Mutational burden in each cancer is ordered by percentile rank. Cancer types are ordered on the

horizontal axis based on their median numbers of somatic mutations. Mutational burden of Group 4 tumors is statistically higher than Groups 1–3 ($P=1.6 \times 10^{-3}$, Kruskal Wallis test).

c, Distribution of the number of mutations that are considered clonal per each patient sample (column). A total of 26% of tumors exhibited only clonal point mutations. In the median tumor, 75% of single nucleotide variants were clonal.

d, Cancer cell fraction of all variants in each patient sample (columns) ordered as in (c). Variants are colored according to the classification in the legend.

e, Cancer cell fraction of recurrent oncogenic driver mutations (columns). Variants are colored according to the classification in the legend.

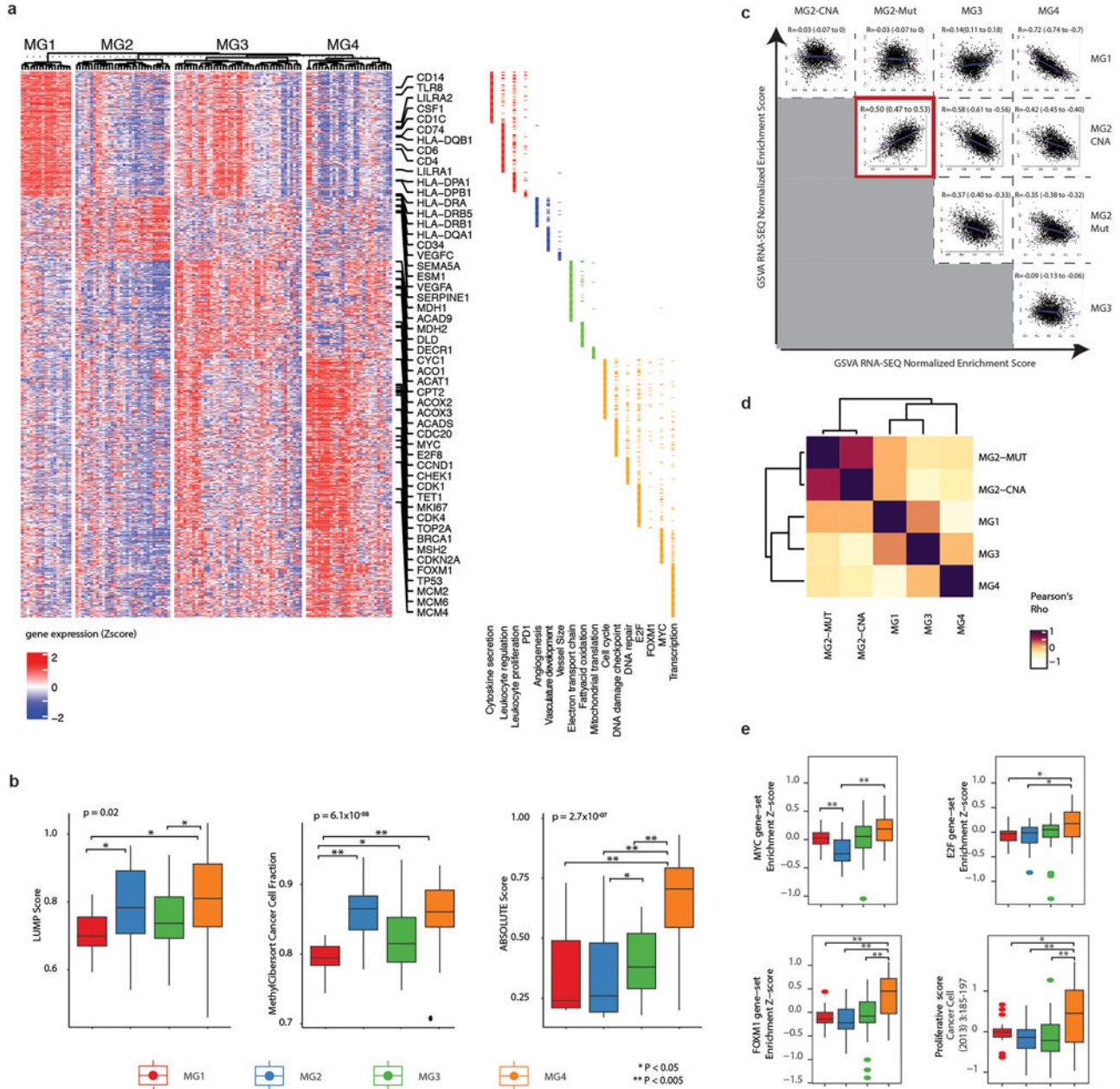


Extended Data Fig. 4: Genomic disruptions differ among molecular groups

a, Genome-wide copy-number alterations computed from whole-exome sequencing data. Arrangements of copy number profile are matched to the samples from mutation plot above. Only mutations that are relevant to discussion in text are shown.

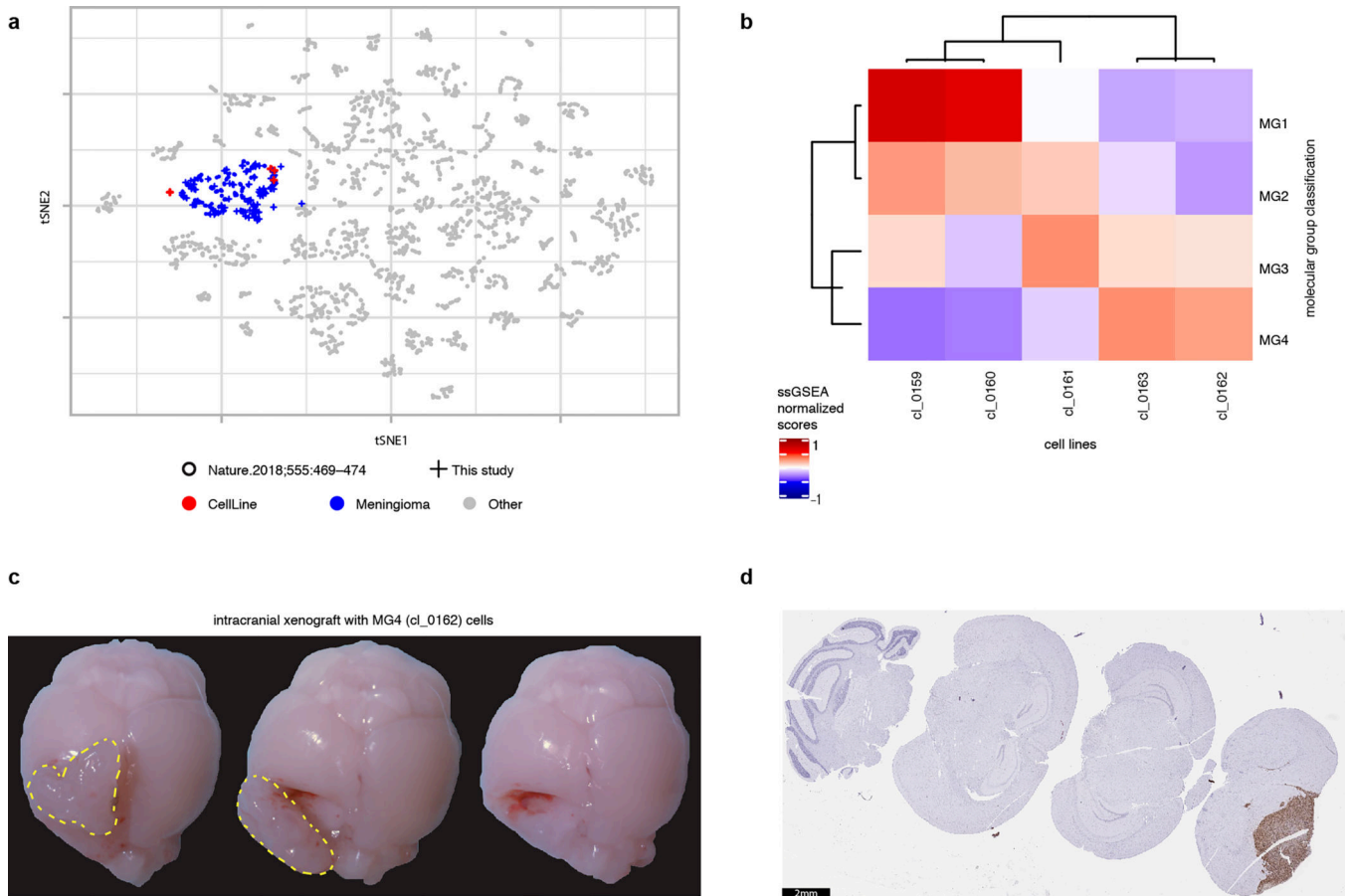
b, Boxplots showing the mRNA expression of NF2 stratified by molecular group. Each dot is a sample. Samples are colored by NF2 mutation status and shapes are according to NF2 deletion status by CNA. Some MG3 and MG4 meningiomas that are NF2 wildtype show silencing of NF2 expression.

c, Boxplots comparing the mean methylation level of NF2 wildtype MG3 and MG4 meningiomas with high versus low NF2 expression using all probes (left), those mapping to the promoter region (middle), and those mapping to the gene body (right).
d, Circos plot showing the landscape of interchromosomal gene rearrangements detected using a stringent threshold for conservative estimation of fusion events (unique spanning reads > 25) in each molecular group. Total number of interchromosomal fusion in MG1, MG2, MG3 and MG4 are 2, 7, 18, and 23, respectively.



Extended Data Fig. 5: Gene expression profiles of molecular groups

- a,** Hierarchical clustering of the expression of genes from select pathways identified in Fig. 2a. Selected genes have been labeled. Redundancy of genes to pathways is shown in the side bar.
- b,** Boxplots showing the results for estimates of immune and stromal infiltration by DNA methylation (LUMP score on left and methylCIBERSORT in middle) and somatic DNA alterations (right, ABSOLUTE score).
- c,** Scatterplots comparing normalized enrichment scores between molecular groups using Gene Set Variation Analysis (GSVA). Each dot is a pathway. Shown at the top of each panel are Pearson correlations and associated 95%CI. MG2 tumors were divided into tumors that are driven by CNA (MG2-CNA) and tumors that are driven by mutations (MG2-Mut). Correlations were highest when comparing MG2 tumors driven by CNA to MG2 tumors driven but mutations (red box).
- d,** Hierarchical clustering of normalized enrichment scores from (c) identifies MG2-CNA and MG2-Mut tumors as one coherent group.
- e,** Boxplots comparing the activation of molecular proliferative signatures between MGs. Statistical significance is denoted by asterisks.



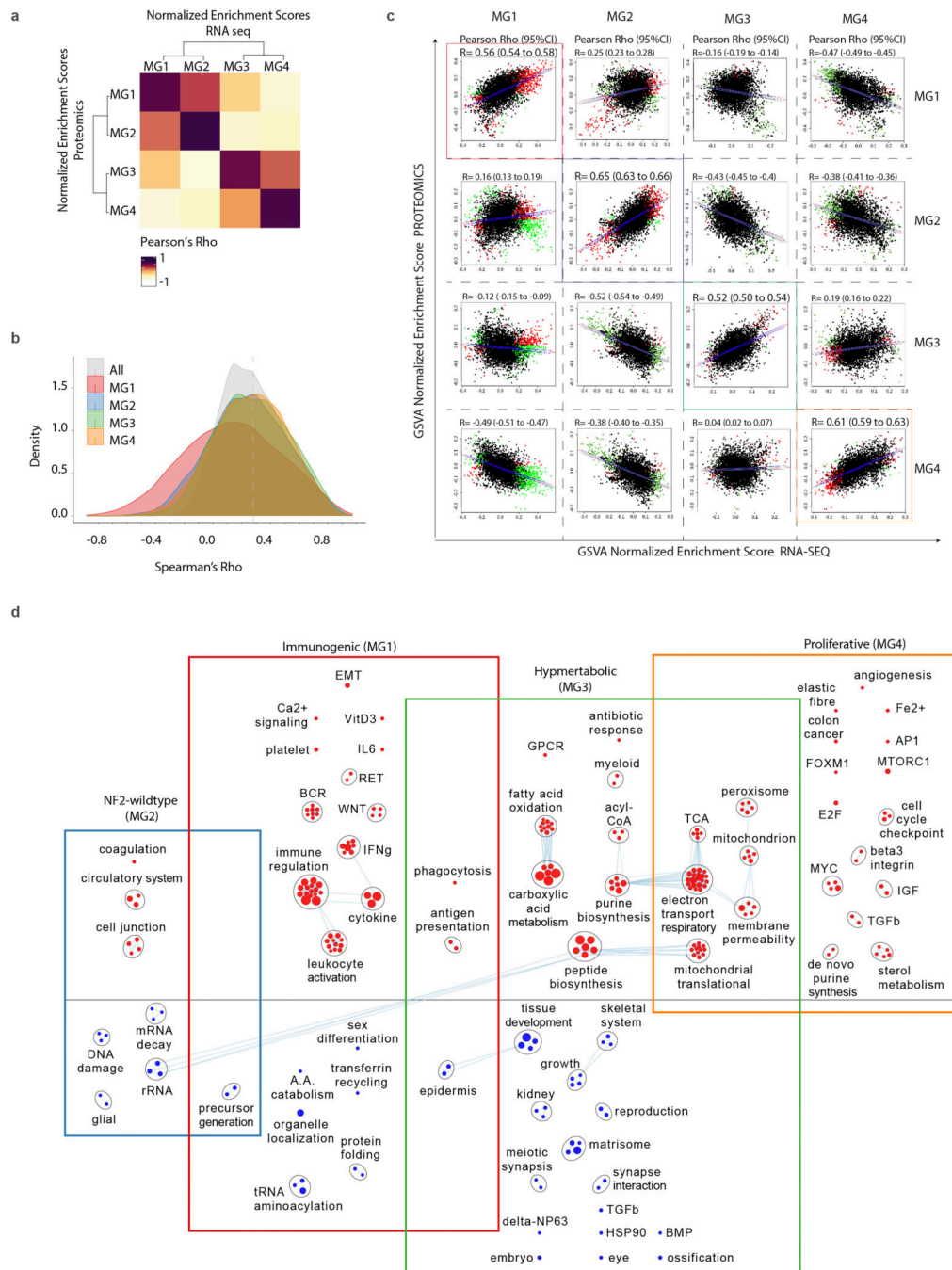
Extended Data Fig. 6: Molecular characterization of patient derived cell lines

a, t-distributed Stochastic Neighbor Embedding (tSNE) plot of genome-wide DNA methylation profiles of patient derived cell lines (red), to meningiomas (blue), and 2798 previously published tumors from 40 other brain tumor types⁵⁸.

b, Heatmap showing results of single-sample Gene-Set Enrichment Analysis (ssGSEA) using mRNA data from cell lines. Each cell line was assigned a score for Molecular Groups 1, 2, 3 and 4 using gene-expression based signatures from the discovery cohort. MG designation was determined by highest scores from ssGSEA assignments.

c, Gross morphological images of a representative MG4-xenografted mice. Extra axial tumor is outlined in dashed yellow lines. Compression on adjacent neural structures is evident after partial (middle panel) and complete (right panel) separation of meningioma from brain.

d, Serial sections and immunostaining for MCM2 in representative MG4-xenografted mice. Scale bar is 2mm. Small areas of tumor that have invaded the brain can be seen staining for MCM2.



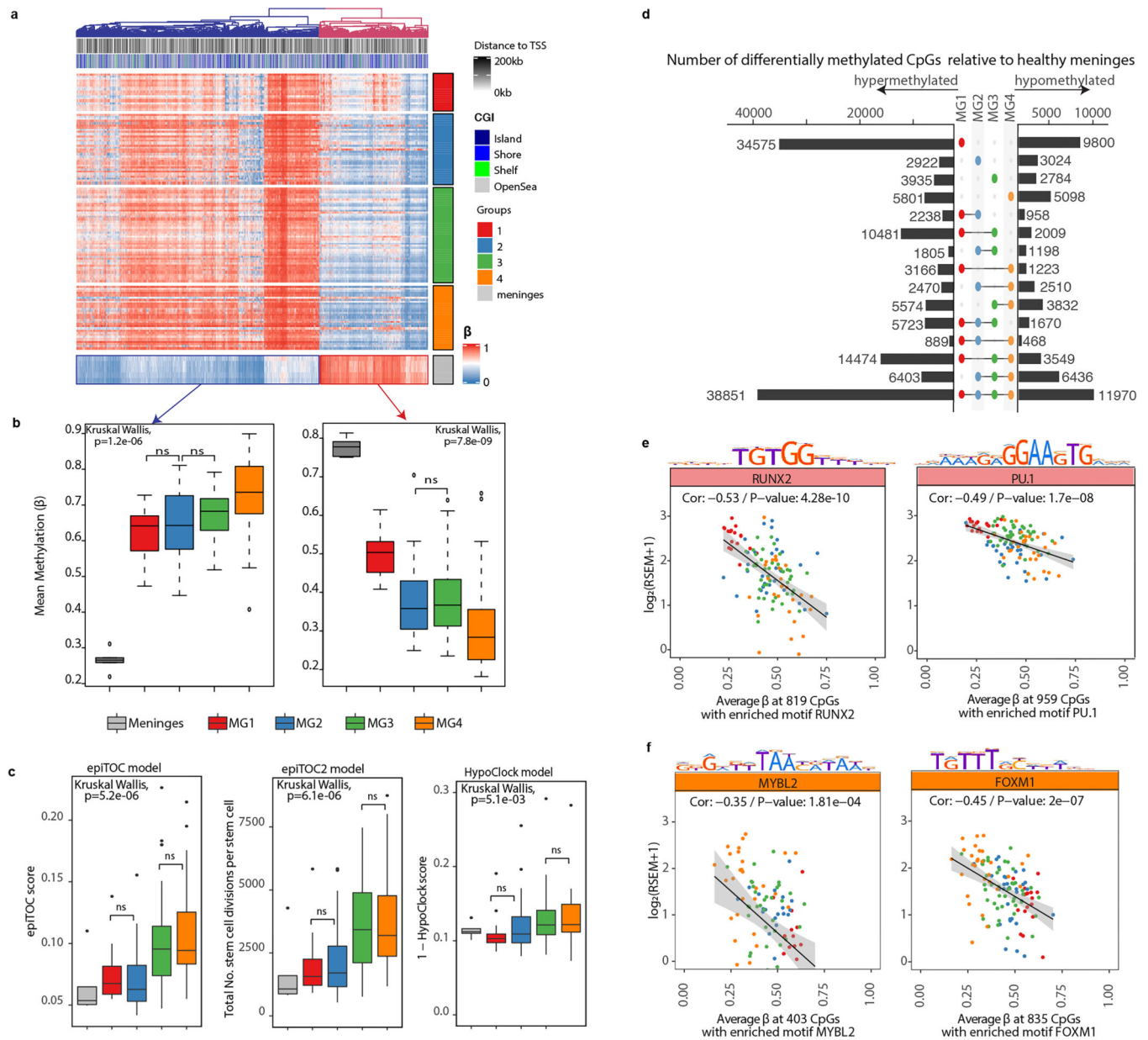
Extended Data Fig. 7: Proteomic and gene expression data converge to similar biology driving each molecular group

a, Hierarchical clustering of normalized enrichment scores obtained by Gene-Set Variation Analysis (GSVA) using proteomic data (rows) and mRNA data (columns).

b, Distribution of correlation of mRNA expression to protein abundance in all samples (grey), MG1 meningiomas (red), MG2 meningiomas (blue), MG3 meningiomas (green) and MG4 meningiomas (orange). Vertical line indicates overall median correlation across all samples (Spearman's $r = 0.279$, 95%CI 0.273–0.284).

c, Scatterplots comparing normalized enrichment scores by GSVA using gene expression (x-axis) and protein abundance (y-axis) stratified by MG classifications. Each dot represents a pathway. Pathways that are statistically significant and concordant by protein and mRNA data are colored green while those that are discordant are colored red. Pearson correlations and 95% confidence intervals are indicated at the top of each panel.

d, Network of activated gene circuits by proteome data in N=96 samples. Protein groups were ranked for each subtype by degree of differential expression. Gene-set enrichment analysis was performed on the ranked gene lists and enriched pathways are visualized using the EnrichmentMap plugin in Cytoscape App. Nodes represent pathways and edges represent shared genes between pathways. Pathways above horizontal line are up-regulated (red nodes) in each molecular group while pathways below horizontal line are down-regulated (blue nodes) in each molecular group.



Extended Data Fig. 8: Differences in genome-wide methylation across meningioma groups

a, Hierarchical clustering of highly differentially methylated CpGs (absolute $\Delta\beta > 0.35$, FDR < 0.05) between all meningiomas and healthy meninges. Annotations of MGs are on the right side of heatmap.

b, Boxplots showing the distribution of β values for probes in (a) that are hypermethylated in healthy meninges (left) and hypomethylated in healthy meninges (right). Pairwise comparisons in each boxplot are statistically significant ($p < 0.05$), unless explicitly stated otherwise (ns, not significant).

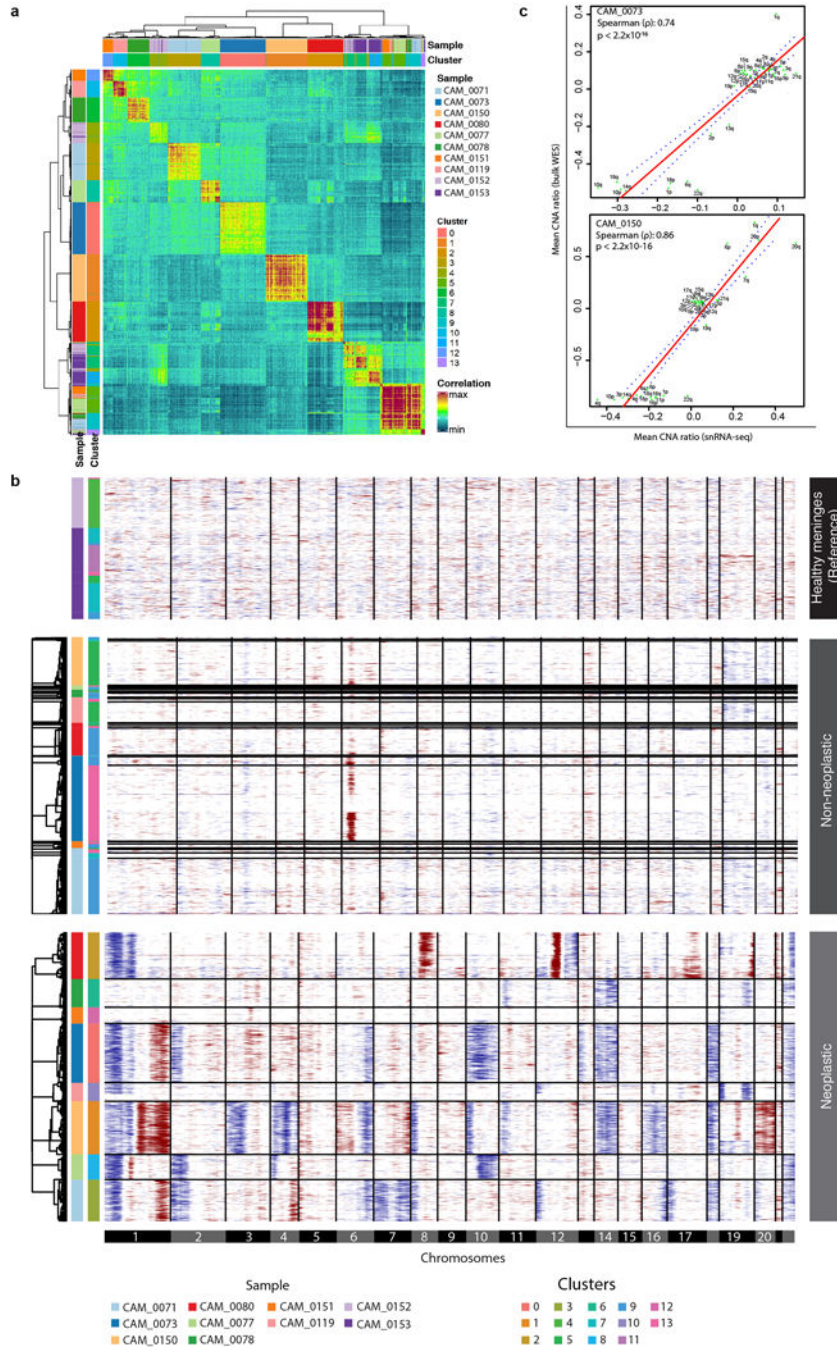
c, Boxplots showing the distribution of using epigenetic mitotic clocks with epiTOC model (left), epiTOC2 model (middle), and HypoClock model (right). Pairwise comparisons in

each boxplot are statistically significant ($p < 0.05$), unless explicitly stated otherwise (ns, not significant).

d, Number of unique and overlapping probes that are differentially methylated (absolute $\Delta\beta > 0.1$, FDR < 0.05) when comparing each MG group to healthy meninges. Bar plot on the left indicates the total number of probes that were hypomethylated in each comparison, and barplot on the right indicates total number of probes that were hypermethylated in each comparison.

e, Scatterplots comparing master regulator transcription factor expression with average β at sites enriched for the motif of that transcription factor. Samples are colored according to molecular group. Pearson correlation with 95% confidence intervals are reported.

Hypomethylation at motifs of immunological-lineage-specific transcription factors such as PU.1, RUNX1/2 and IRF5/8 were enriched in immunogenic (MG1) tumors ($P = 1.05 \times 10^{-8}$, hypergeometric test) and associated with enhancer hypomethylation. Similarly, master regulators of cell proliferation such as MYBL2, LHX4, and FOXM1 were hypomethylated in proliferative (MG4) tumors and associated with increased abundance of these transcription factors ($P = 1.24 \times 10^{-3}$, hypergeometric test).



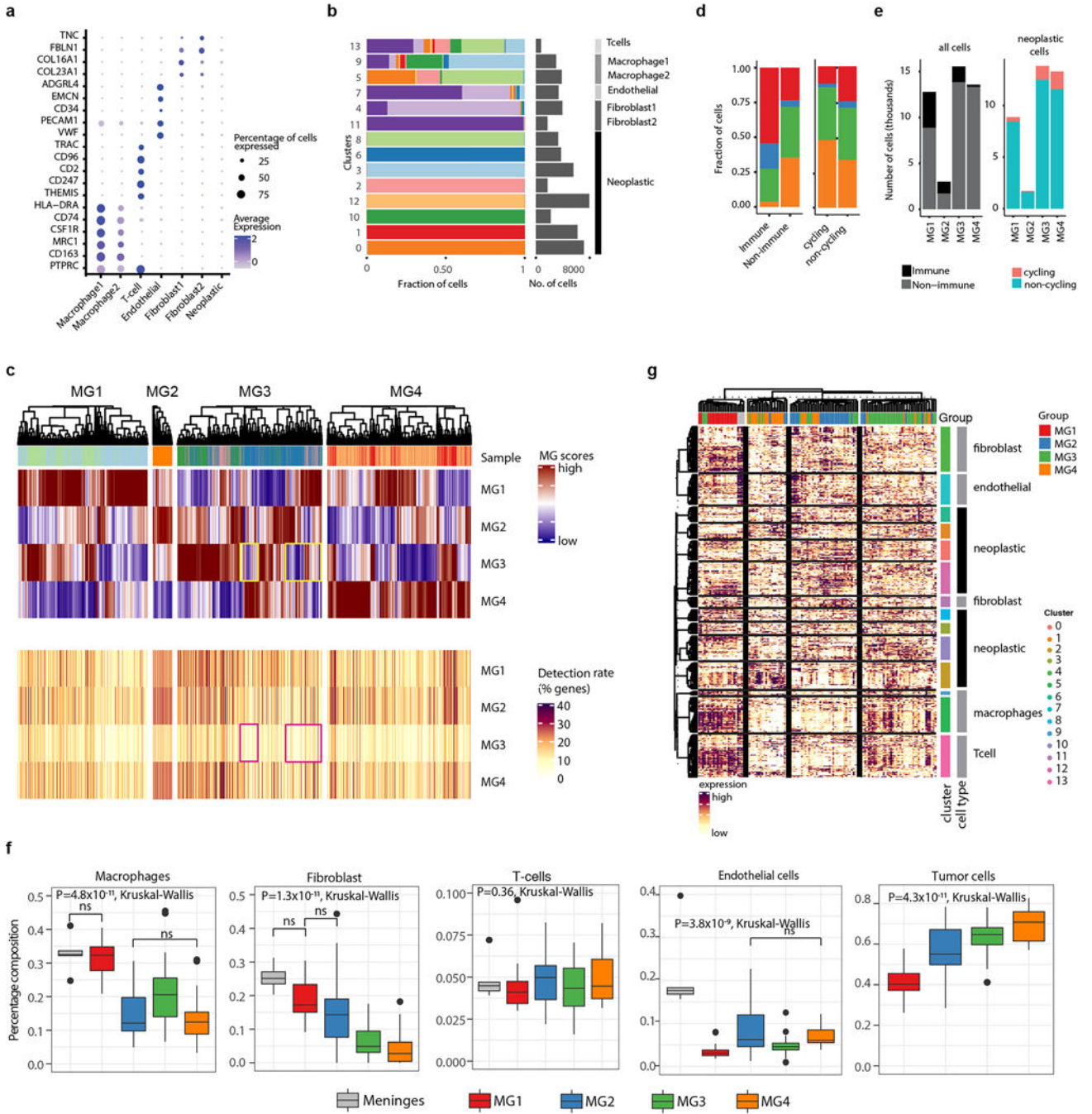
Extended Data Fig. 9: Meningiomas show low within patient variation of expression and copy number profile

a, Pairwise correlations of expression profiles of all cells ordered by hierarchical clustering. Each cell is annotated to tumor of origin from Fig. 4a and cluster assignments from Fig. 4b at top and side bars.

b, Inferred genome-wide copy number variations of single nuclei of healthy meninges (reference, top panel), immune cells (middle panel), and neoplastic cells (bottom panel). Sample and cluster annotation are shown on the left. The copy number plot of these tumors

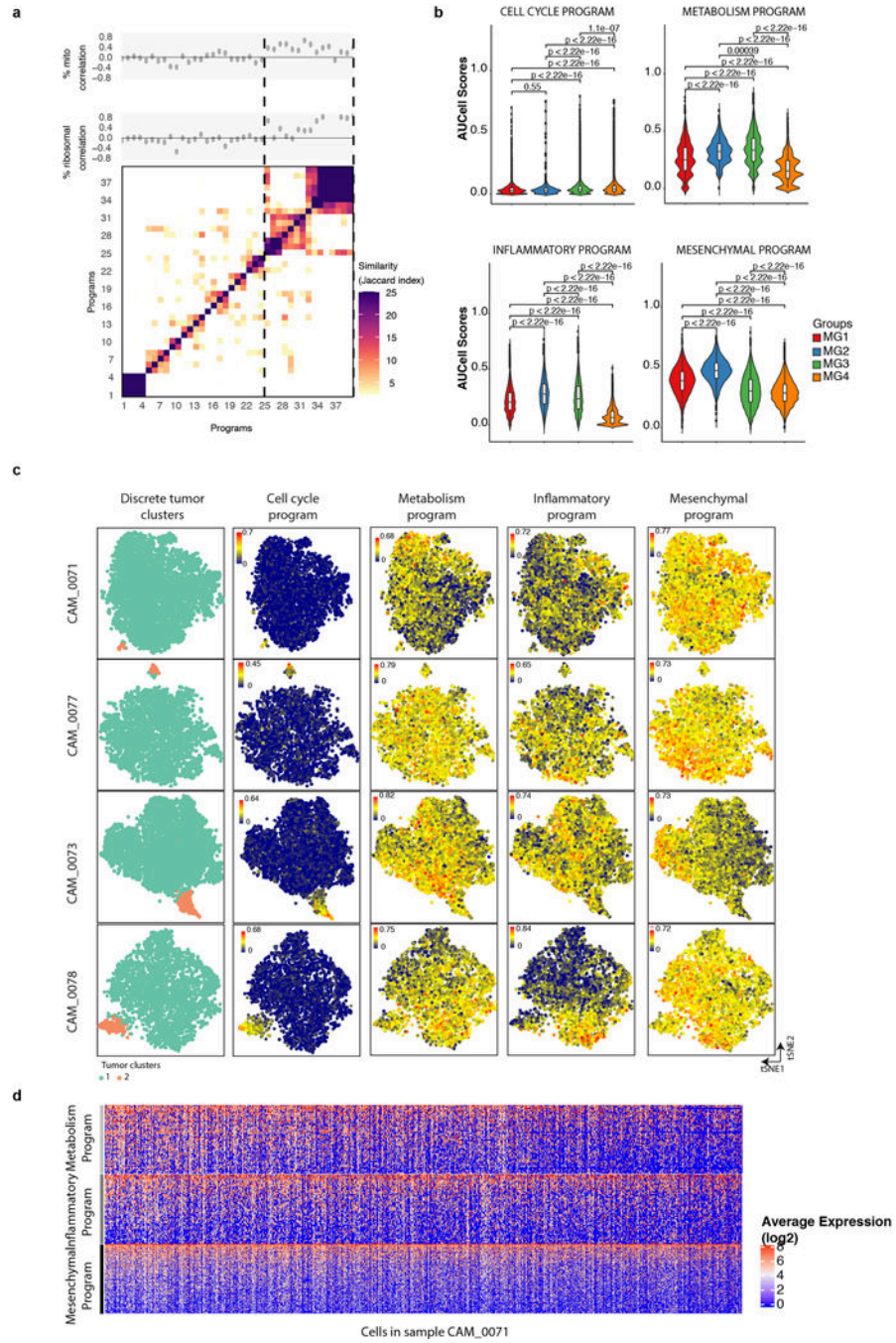
are homogenous and subclones of cells within tumors with distinct copy number profiles are not readily identifiable. Annotation to patient of origin and cluster on the left of each heatmap.

c. Scatterplots showing the relationship between arm-level CNA inferred by snRNA-seq (x-axis) to matched CNA by bulk whole exome sequencing (y-axis). Two representative samples are shown.



Extended Data Fig. 10: The transcriptome of MGs is shaped by the expression profiles from both neoplastic and non-neoplastic cells

- a**, Bubble plot showing the expression of lineage specific markers for distinct cell types.
- b**, Stacked barplot showing the relationship of samples to clusters. Samples are colored by patient of origin as in Fig. 4a. Barplot to the right shows the number of cells within each cluster.
- c**, The top heatmap shows hierarchical clustering results of single cells by MG scores. Each cell was scored for the bulk signature of each MGs and scores were compared to a permuted random gene set. Shown are cells with at least one score with $FDR < 0.2$. Scores were scaled such that the sum of all scores for each cell is equal to one. Below is a matched heatmap showing the number of genes detected for each MG signature in each cell. In a subset of cells, low scores are associated with low detection rate of genes (yellow and pink boxes).
- d**, Stacked barplot showing the distribution of immune versus non-immune cells across molecular groups (left) and cycling versus non-cycling neoplastic cells across molecular groups (right) to clusters. Samples are colored by molecular group of tumor as in Fig. 4d.
- e**, Barplot showing the total number of cells that are immune versus non-immune (left) and cycling versus non-cycling (right) by MG status of tumor of origin
- f**, Boxplots comparing the cell type composition of bulk RNA seq samples after deconvolution using single cell RNA-seq signatures. Pairwise comparisons in each boxplot are statistically significant ($p < 0.05$), unless explicitly stated otherwise (ns, not significant).
- g**, Heatmap showing the expression of marker genes for single cell clusters (determined by CIBERSORTx) in bulk RNA seq data. Each column represents one tumor. Rows are designated marker genes for each cluster. Tumors were partitioned into 4 partitions by consensus k-means clustering with samples and gene sets clustered by hierarchical clustering using Pearson distance metric.

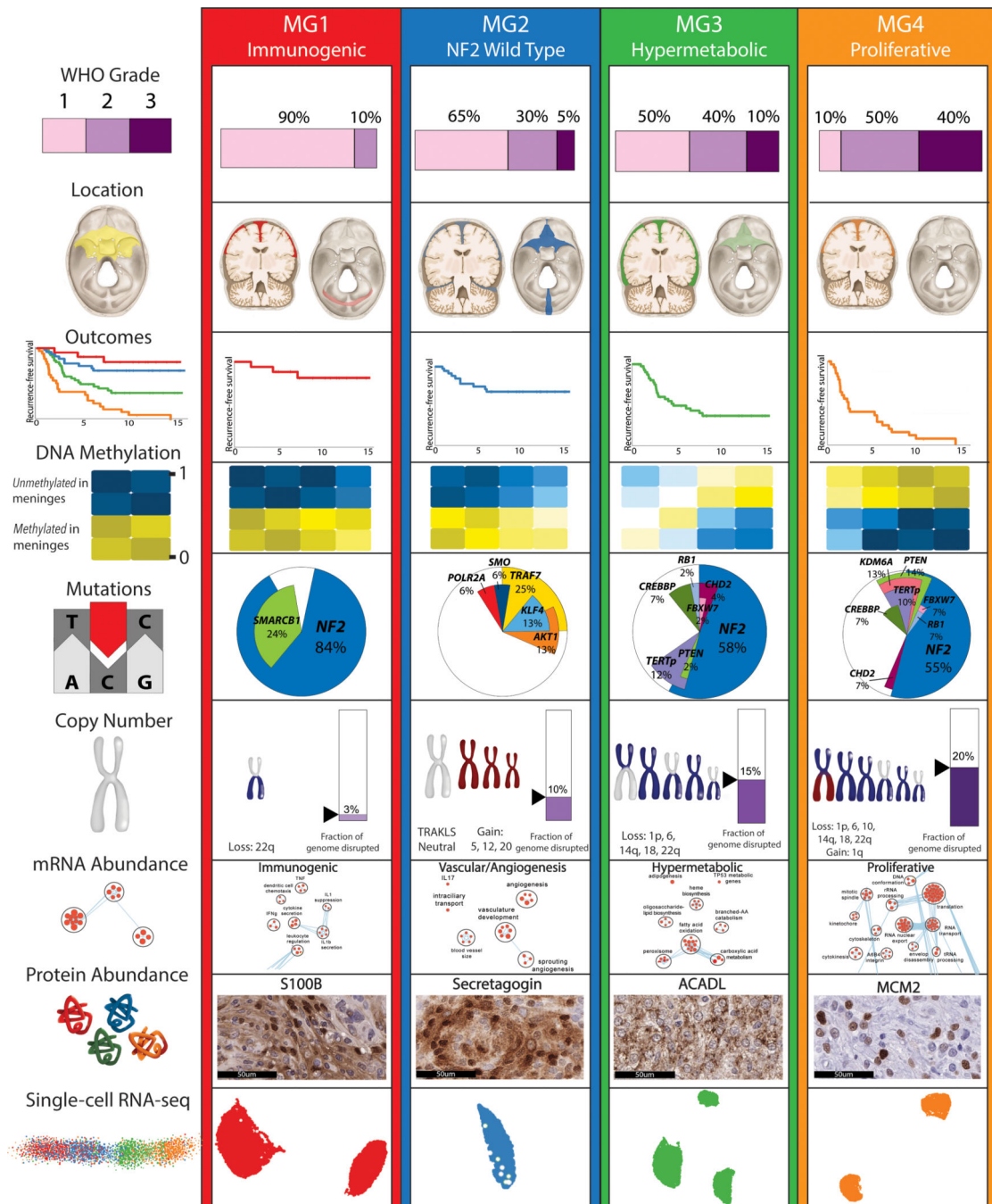


Extended Data Fig. 11: Discrete and continuous patterns of variability can be identified in meningioma

a. Hierarchical clustering of similarities between NMF programs. Top panel indicates Pearson correlations between number of mitochondrial and ribosomal genes detected with NMF scores for each program. A cluster of programs (dashed lines) showed positive correlation with the expression of mitochondrial and ribosomal genes (confirmed by manual inspection). These programs were considered to be reflective of technical artifacts and not included in subsequent analyses.

b, Violin plots showing the distribution of activation scores for NMF programs across MGs.
c, Side-by-side tSNEs showing the relationship of discrete clustering results with activation scores of each NMF program. Shown are four representative samples. Activation scores of cell cycle program are closely associated with discrete clusters, whereas scores of metabolism, inflammatory, and mesenchymal program are not associated with discrete clusters.

d, Heatmap showing the average expression of genes defining NMF programs (annotated to left) in representative sample CAM_0071. Cells are ranked and ordered according to the activation score of the metabolism program. There is a continuous pattern of gene expression variability in these programs.



Extended Data Fig. 12. Graphical summary of findings.

Shown is a schematic representation that summarizes the major molecular findings and conclusions of our study: unsupervised consensus clustering combining DNA copy number, DNA methylation, and mRNA sequencing data revealed four robust groups of tumors with prototypical biology and distinct clinical outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments and Funding:

F.N. is supported by the Canadian Institute of Health Research (CIHR) Vanier Scholarship, AANS/CNS Section on Tumors & NREF Research Fellowship Grant, and Hold'em for Life Oncology Fellowship. G.Z is funded by a CIHR Project Grant Award (159452), and the Brain Tumor Charity UK Quest for Cures Grant (GN-000430). P.C.B. was supported by NIH/NCI under awards P30CA016042, U24CA248265 and P50CA211015.

Competing Interests:

D.D.D.C., and A.C. are listed as inventors on patents filed that are unrelated to this project. D.D.D.C. received research funding from Pfizer and Nektar therapeutics not related to this project.

References

1. Ostrom QT et al. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012–2016. *Neuro. Oncol* 21, v1–v100 (2019). [PubMed: 31675094]
2. Goldbrunner R. et al. EANO guidelines for the diagnosis and treatment of meningiomas. *Lancet. Oncol* 17, e383–91 (2016). [PubMed: 27599143]
3. Sahm F. et al. DNA methylation-based classification and grading system for meningioma: a multicentre, retrospective analysis. *Lancet Oncol.* 18, 682–694 (2017). [PubMed: 28314689]
4. Clark VE et al. Genomic analysis of non-NF2 meningiomas reveals mutations in TRAF7, KLF4, AKT1, and SMO. *Science* (80-.) 339, 1077–1080 (2013).
5. Brastianos PK et al. Genomic sequencing of meningiomas identifies oncogenic SMO and AKT1 mutations. *Nat. Genet* 45, 285–289 (2013). [PubMed: 23334667]
6. Clark VE et al. Recurrent somatic mutations in POLR2A define a distinct subset of meningiomas. *Nat. Genet* 48, 1253–1259 (2016). [PubMed: 27548314]
7. Harmancl AS et al. Integrated genomic analyses of de novo pathways underlying atypical meningiomas. *Nat. Commun* 8, 14433 (2017). [PubMed: 28195122]
8. Nassiri F. et al. DNA methylation profiling to predict recurrence risk in meningioma: development and validation of a nomogram to optimize clinical management. *Neuro. Oncol* 21, 901–910 (2019). [PubMed: 31158293]
9. Hoadley KA et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* 173, 291–304.e6 (2018). [PubMed: 29625048]
10. Koboldt DC et al. Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70 (2012). [PubMed: 23000897]
11. Cancer Genome Atlas Research Network et al. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med* 372, 2481–98 (2015). [PubMed: 26061751]
12. Hoadley KA et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 158, 929–944 (2014). [PubMed: 25109877]
13. Reshef DN et al. Detecting Novel Associations in Large Data Sets. *Science* (80-.) 334, 1518–1524 (2011).
14. Zhou W. et al. DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet* 50, 591–602 (2018). [PubMed: 29610480]
15. Teschendorff AE A comparison of epigenetic mitotic-like clocks for cancer risk prediction. *Genome Med.* 12, 56 (2020). [PubMed: 32580750]
16. Yang Z. et al. Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biol.* 17, 205 (2016). [PubMed: 27716309]
17. Silva TC et al. ELMER v.2: an R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *Bioinformatics* 35, 1974–1977 (2019). [PubMed: 30364927]

References (Methods only)

18. Aryee MJ et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30, 1363–1369 (2014). [PubMed: 24478339]
19. Ritchie ME et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47–e47 (2015). [PubMed: 25605792]
20. Li H. & Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760 (2009). [PubMed: 19451168]
21. McKenna A. et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303 (2010). [PubMed: 20644199]
22. Cibulskis K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol* 31, 213–219 (2013). [PubMed: 23396013]
23. Saunders CT et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 28, 1811–1817 (2012). [PubMed: 22581179]
24. Lek M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291 (2016). [PubMed: 27535533]
25. Lawrence MS et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218 (2013). [PubMed: 23770567]
26. McLaren W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 17, 122 (2016). [PubMed: 27268795]
27. Chakravarty D. et al. OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol* 1–16 (2017) doi:10.1200/PO.17.00011.
28. Chang MT et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat. Biotechnol* 34, 155–163 (2016). [PubMed: 26619011]
29. Dong C. et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet* 24, 2125–2137 (2015). [PubMed: 25552646]
30. Talevich E, Shain AH, Botton T. & Bastian BC CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLOS Comput. Biol* 12, e1004873 (2016).
31. Cerami E. et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discov.* 2, 401–404 (2012). [PubMed: 22588877]
32. Dobin A. et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* (2013) doi:10.1093/bioinformatics/bts635.
33. Li H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009). [PubMed: 19505943]
34. Aibar S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* 14, 1083–1086 (2017). [PubMed: 28991892]
35. Robinson MD, McCarthy DJ & Smyth GK edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140 (2010). [PubMed: 19910308]
36. Wilkerson MD & Hayes DN ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26, 1572–1573 (2010). [PubMed: 20427518]
37. D'Entrop SC, Wedge DC & Van Loo P. Principles of Reconstructing the Subclonal Architecture of Cancers. *Cold Spring Harb. Perspect. Med* 7, a026625 (2017).
38. Aran D, Sirota M. & Butte AJ Systematic pan-cancer analysis of tumour purity. *Nat. Commun* 6, 8971 (2015). [PubMed: 26634437]
39. Reimand J. et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc* 14, 482–517 (2019). [PubMed: 30664679]
40. Langmead B, Trapnell C, Pop M. & Salzberg SL Bowtie: An ultrafast memory-efficient short read aligner. [<http://bowtie.cbcb.umd.edu/>]. *Genome Biol.* (2009) doi:10.1186/gb-2009-10-3-r25.
41. Langmead B. & Salzberg S. Bowtie2. *Nat. Methods* (2013) doi:10.1038/nmeth.1923.Fast.

42. Kent WJ BLAT--The BLAST-Like Alignment Tool. *Genome Res.* (2002) doi:10.1101/gr.229202.
43. Lambert SA et al. The Human Transcription Factors. *Cell* 172, 650–665 (2018). [PubMed: 29425488]
44. Wingender E, Schoeps T, Haubrock M, Krull M. & Dönitz J. TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res.* 46, D343–D347 (2018). [PubMed: 29087517]
45. Cox J. & Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol* 26, 1367–1372 (2008). [PubMed: 19029910]
46. A global reference for human genetic variation. *Nature* 526, 68–74 (2015). [PubMed: 26432245]
47. Kang HM et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol* 36, 89–94 (2018). [PubMed: 29227470]
48. Lun AT, Bach K. & Marioni JC Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* 17, 75 (2016). [PubMed: 27122128]
49. Lun ATL, McCarthy DJ & Marioni JC A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* 5, 2122 (2016). [PubMed: 27909575]
50. Butler A, Hoffman P, Smibert P, Papalexi E. & Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol* 36, 411–420 (2018). [PubMed: 29608179]
51. Tickle TI, Georgescu C, Brown M. & Haas B. inferCNV of the Trinity CTAT Project. (2019).
52. Tirosh I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* (80-.) 352, 189–196 (2016).
53. Azizi E. et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell* 174, 1293–1308.e36 (2018). [PubMed: 29961579]
54. Jerby-Arnon L. et al. Opposing immune and genetic mechanisms shape oncogenic programs in synovial sarcoma. *Nat. Med* 27, 289–300 (2021). [PubMed: 33495604]
55. Kinker GS et al. Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat. Genet* 52, 1208–1218 (2020). [PubMed: 33128048]
56. Izar B. et al. A single-cell landscape of high-grade serous ovarian cancer. *Nat. Med* 26, 1271–1279 (2020). [PubMed: 32572264]
57. Newman AM et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol* 37, 773–782 (2019). [PubMed: 31061481]
58. Capper D. et al. DNA methylation-based classification of central nervous system tumours. *Nature* 555, 469–474 (2018). [PubMed: 29539639]

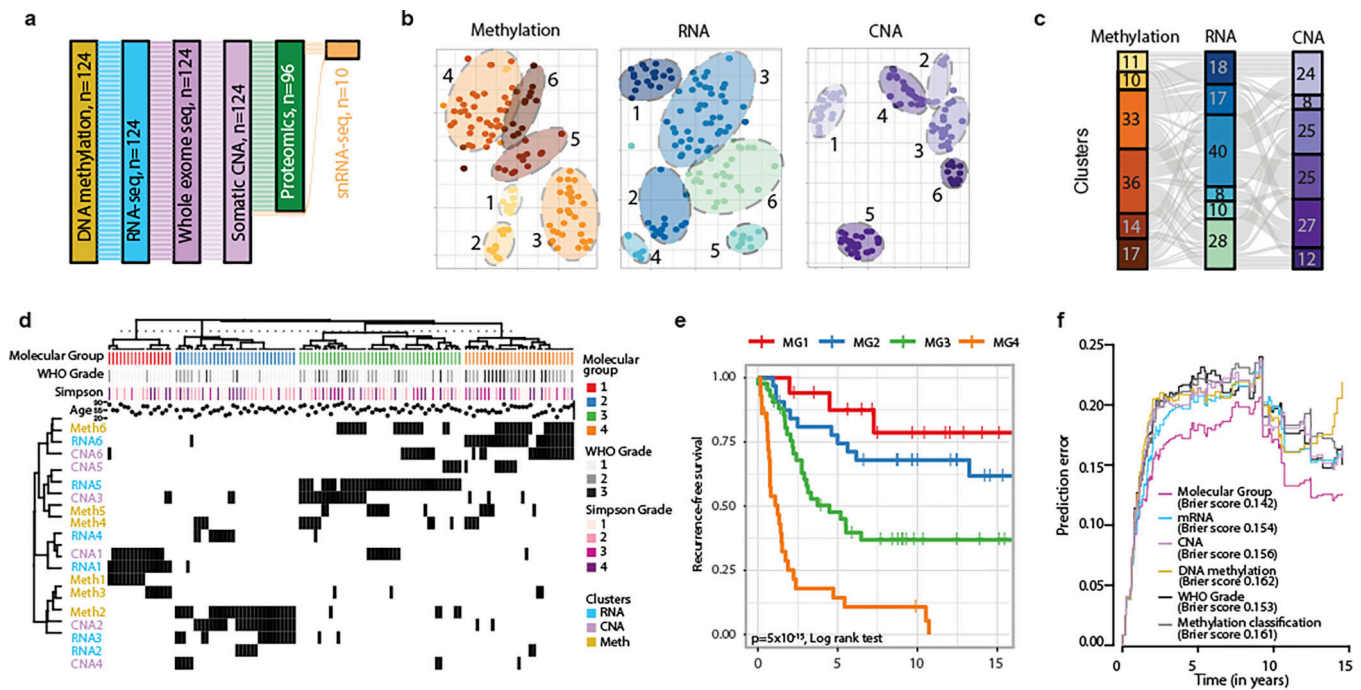


Figure 1. Integrative multiplatform analysis reveals 4 novel molecular groups of meningioma

a, Flow diagram showing relationship of molecular datasets in this study: whole-exome sequencing, DNA methylation and mRNA sequencing (n=124), proteomics (n=96), and single cell data (n=8). A total of 121 samples were used for discovery on bulk analyses, with an additional 3 samples assembled specifically for single cell analyses.

b, t-distributed stochastic neighbor embedding (tSNE) reduction of individual platform data with annotated unsupervised cluster assignments for each individual platform.

c, Alluvial plot showing relationships of unsupervised cluster assignments from individual platforms analyses using DNA methylation, RNA sequencing, and copy number data. Width of the nodes and edges are proportional to the number of samples.

d, Multiplatform higher order integration of genetic, epigenetic and transcriptomic data by Cluster-of-Cluster assignments. Cluster assignments for each independent platform (rows) are shown for each sample (columns). Membership for a given cluster is noted by a black tick. Annotation for clinical factors: WHO grade and the extent of resection (Simpson grade) are shown above the matrix.

e, Kaplan Meier estimates of recurrence-free survival of patients according to molecular groups.

f, Brier prediction curve for recurrence-free survival comparing classification by molecular groups to WHO grade, DNA methylation cluster assignments by the DKFZ, and cluster assignments by the individual data types in this study. The integrated Brier score is shown for each datatype.

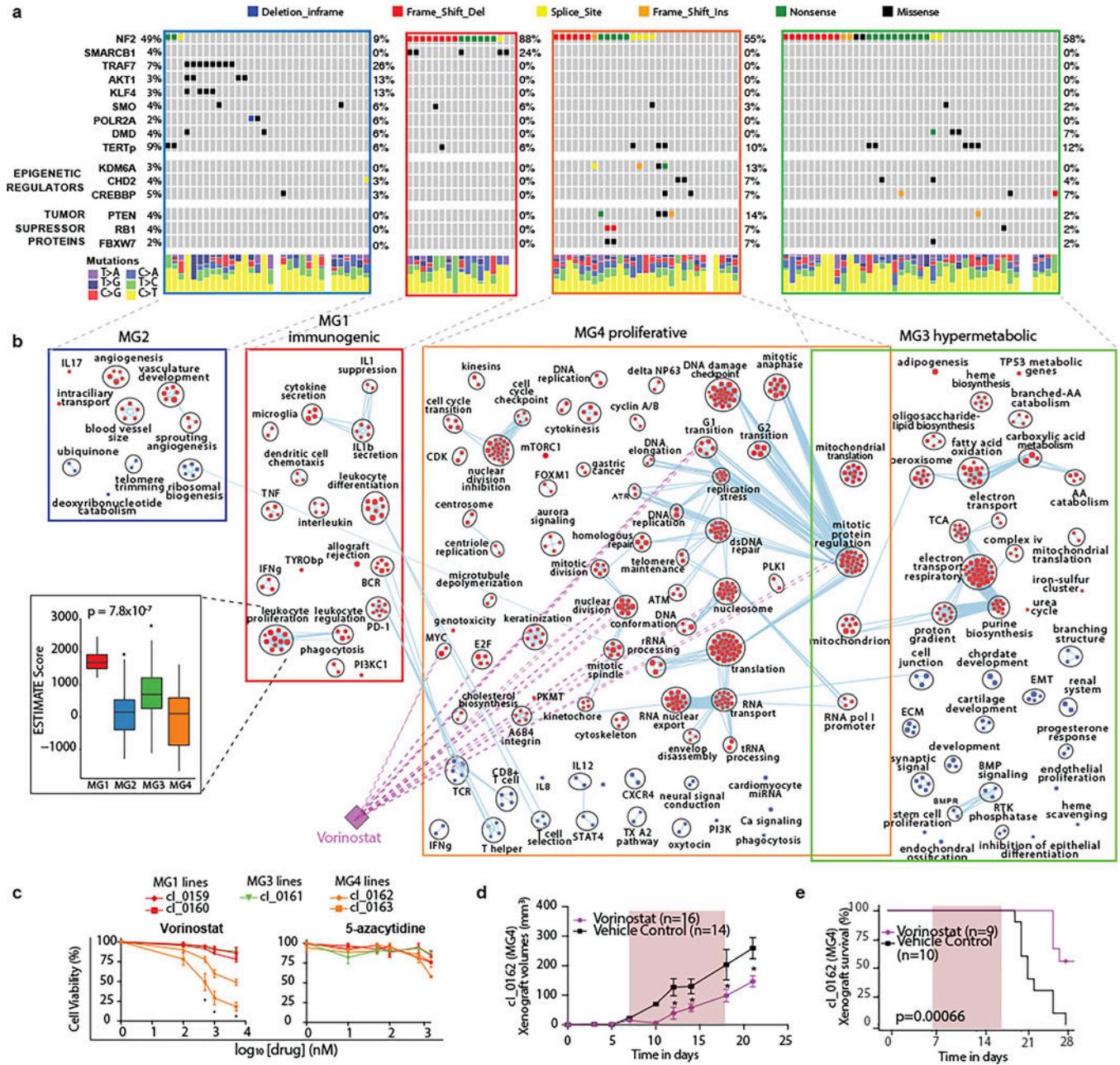


Figure 2: Molecular groups are distinguished by prototypical biology that inform on novel therapeutics

a, Oncoprint showing the recurrent somatic mutations identified in samples in this study. Novel and recurrent mutations in epigenetic regulators and tumor suppressor genes are detected. Colors in oncoprint represent different types of somatic alterations. The relative proportions of the six different possible base-pair substitutions across all variants in each sample are shown in the bottom of the panel.

b, Network of distinguishing gene circuits for each molecular group by mRNA abundance. Nodes represent pathways and edges represent shared genes between pathways. Nodes colored in red are up-regulated pathways while those colored in blue are downregulated

pathways in each molecular group. Light purple edges represent the pathways that are targeted by Vorinostat. Inset shows a boxplot comparing inferred immune cell infiltrates by ESTIMATE between MGs.

c, Results of cell viability assay testing the efficacy of Vorinostat and 5-azacytidine on patient-derived meningioma cell lines that recapitulate specific MGs. Cell lines aligned to MG4 are colored orange, those aligned to MG1 are colored red, and those aligned to MG3 are colored green. * indicates statistical significance by Student's t test at $p < 0.05$. Vorinostat shows a reduction in more than 50% cell viability in MG4 cell lines only, whereas other general treatments such as 5-azacytidine do not show any anti-tumor activity in meningiomas.

d, Tumor volumes of intracranial MG4-xenografted mice measured by serial MRI measurements during treatment with Vorinostat or control. * indicates statistical significance by Student's t test at $p < 0.05$.

e, Kaplan-Meier overall survival distribution of MG4-intracranial xenografts treated with Vorinostat versus control. Statistical significance tested by Log-rank test.

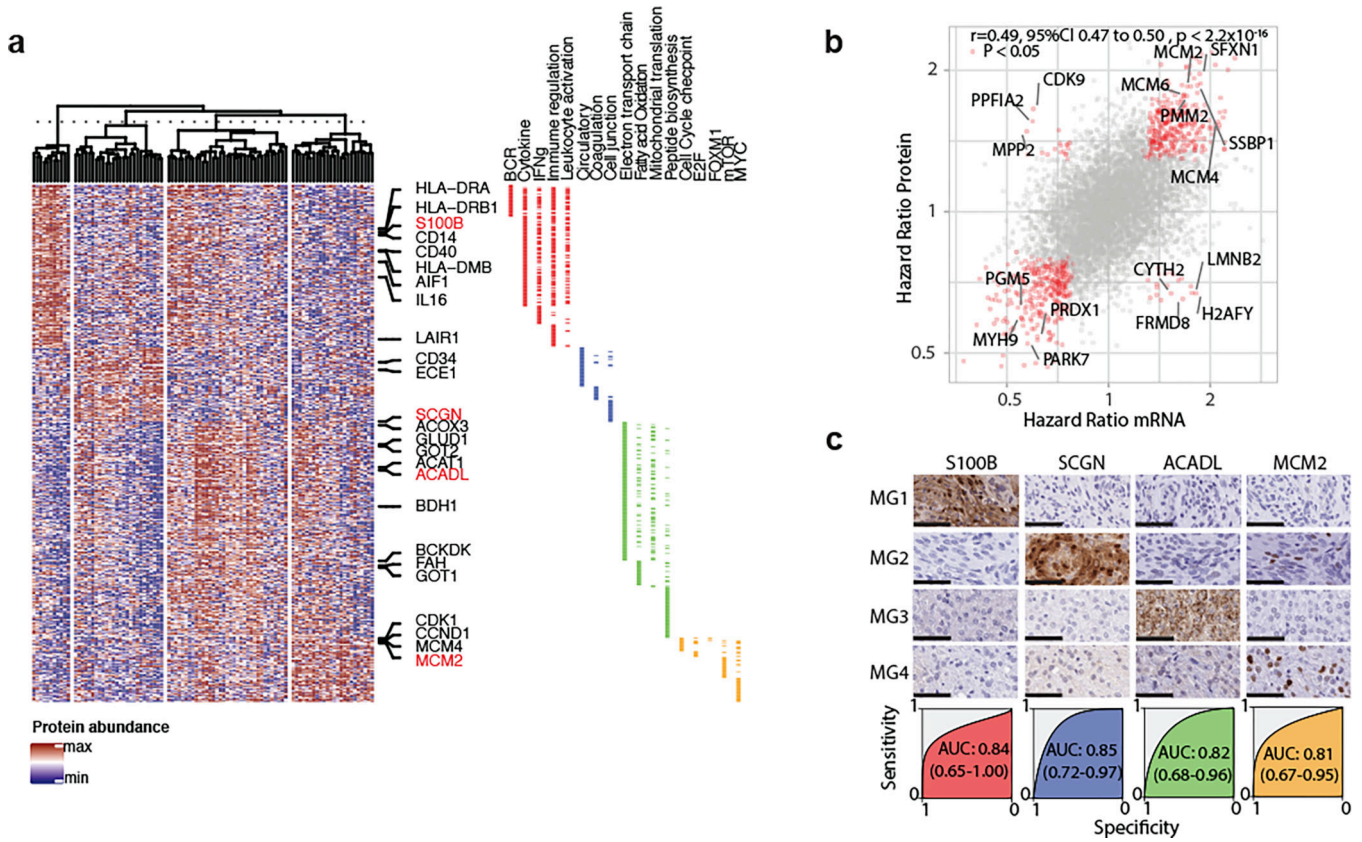


Figure 3: Proteogenomic characterization validates the robustness of MGs and identifies markers that can distinguish MGs by immunohistochemistry

a. Hierarchical clustering of genes from select pathways identified by GSEA. Selected genes have been labeled. Gene annotation to pathway(s) is shown in the side bar.

b. Scatterplot of Hazard Ratios of genes by gene expression (x-axis) and protein abundance (y-axis). Genes with significant associations with outcome are colored in red. Selected genes are labeled. Pearson correlation and its associated 95% confidence interval are shown.

c. Panel showing immunohistochemistry results for group specific markers. Selected are four representative cases (rows). Images shown for each patient are at the same region of the slide for each antibody. Scale bars represent 50um. Each case was subjected to unbiased digital quantitation. Below the panel of representative stains are the Receiver-Operating characteristic (ROC) curve for each antibody with the Area Under the Curve (and 95%CI by Delong method).

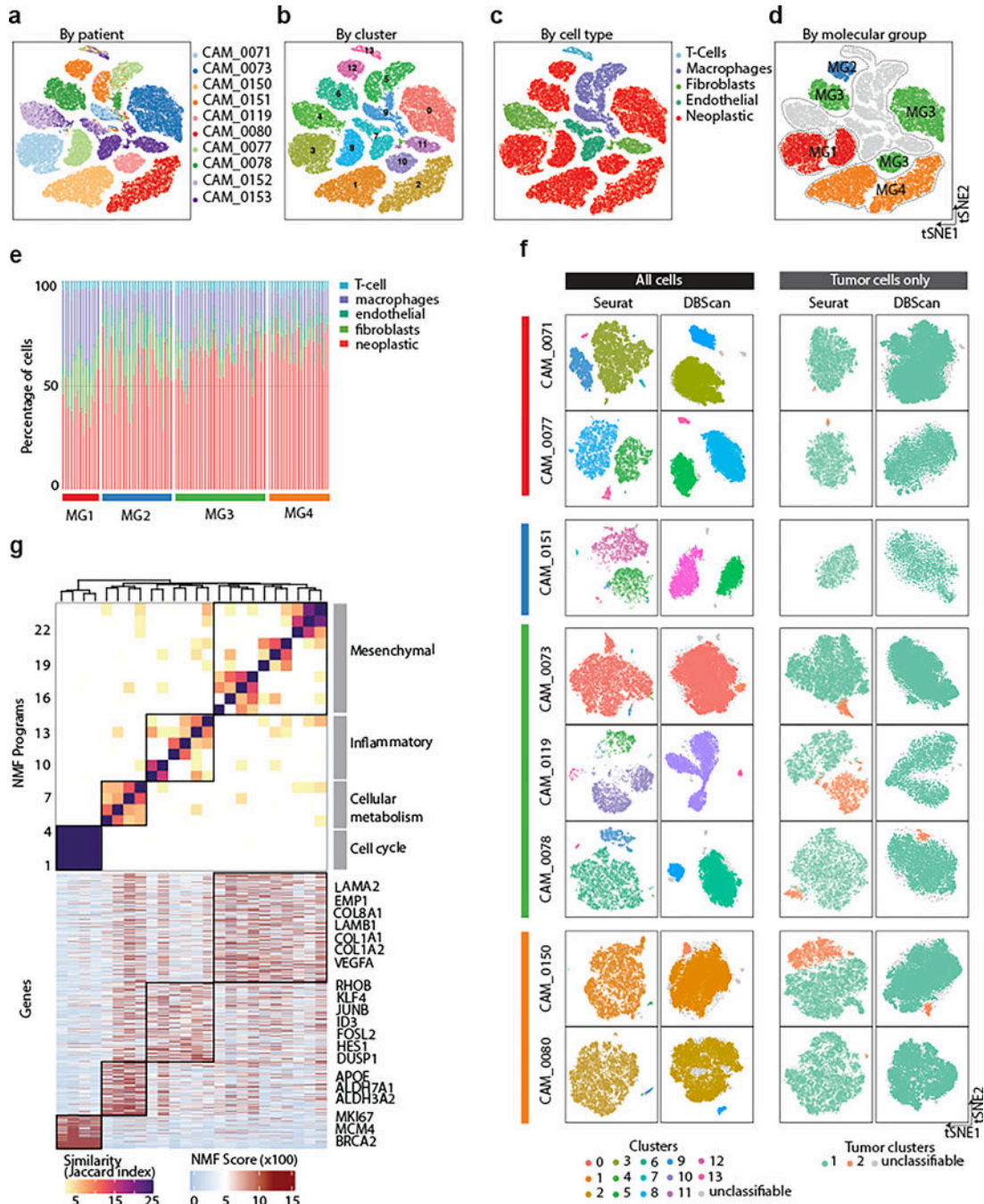


Figure 4: Single cell RNA sequencing of human meningiomas reveals substantial interpatient heterogeneity and subtle within patient variability
a-d, Side-by-side-by-side-by-side t-distributed Stochastic Neighbor Embedding (tSNE) plots of 54,393 nuclei from ten samples, colored by patient of origin (a), cluster number (b), cell type (c), and molecular group of tumor of origin (d).
e, Stacked barplot showing distribution of cell type fractions in bulk RNA-seq data. Samples are grouped according to molecular group. Deconvolution was performed using CIBERSORTx

f, t-distributed Stochastic Neighbor Embedding (tSNE) plot show cluster results for each tumor sample using all cells (left) and tumor cells only (right). Clustering was performed by both Seurat and DBSCAN. Cells that are colored gray by DBScan algorithm did not meet the parameters for clustering (see Methods). For clustering of all cells, the different colors represent the cluster color scheme in (b). For clustering of neoplastic cells, the different colors represent discrete tumor subpopulations. Annotation of patient sample and MG is shown to the left of the tSNEs.

g, Hierarchical clustering of pairwise similarities between non-negative matrix factorization (NMF) programs on the basis of number of shared genes. Four groups of similar programs (meta-programs, black boxes) were identified. The NMF scores for each gene within modules are plotted below as a heatmap. Select genes are labelled. Programs (column) are labelled as in above heatmap.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript